TITLE A STUDY OF THE HISTONE GENES OF *XENOPUS BOREALIS*

AUTHOR Edward Beauchamp Dudliegh Bagenal

INSTITUTION and DATE University of Warwick 1990

CONTENTS

	Page No.
Table of Contents	i
Table of Figures	iii
List of Tables	iv
Acknowledgements	v
Declaration	vi
Abbreviations	vii
Summary	viii
CHAPTER 1 INTRODUCTION	
1A General Introduction	1
1B Histone gene organisation in the following organisms:	5
Stylonychia Mytilis (Protozoan ciliate) and yeast	6
Caenorhadbitis elegans (Nematode)	7
Sea Urchin	8
Drosophila (Fruit Fly)	13
Artemia (Brine Shrimp)	16
Rainbow Trout	18
Newt	19
Xenopus	20
Chicken	28
Mammals	29
'Orphons'	30
1C Histone gene numbers	32
1D Gene Fine structure	35
Conclusion	46
CHAPTER 2 MATERIALS AND METHODS	49
CHAPTER 3 RESULTS: Characterisation of a major cluster clone	
3A General Introduction	74
3B Cloning and initial characterisation of clones	75
3C Characterisation of a major cluster clone	77
3D The sequence of clone XbHW302	77
3E Microinjection of the major cluster clones	84
CHAPTER 4 RESULTS: Genomic organisation of the major cluster	
4A Partial Restriction Digests	88
4B Chromosome Walk	89
4C Clone XbHW102	91
4D Clones XbHW9 and XbHW11	93
4E H4 gene probe	95
4F Pairs or Tandem Repeats	96
4G Total Genomic Digests	100

CONTENTS (Continued) Page No.

CHAPTER 5 RESULTS: Characterisation of the minor cluster.
5A Introduction 103
5B Fine analysis of a minor cluster clone 103
5C Sequence analysis of clone XbHW61 104
5D Microinjection of the minor cluster clones 110
5E H1 genes 113
5F Organisation of the minor cluster within the genome 114

CHAPTER 6 DISCUSSION
6A Introduction 117
6B Introduction to the evolution of histone gene families 118
6C Generation of variant histone gene clusters 121
Coding sequence comparisons 122
Spacer sequence comparisons 125
6D Repetition of new variants 135
Size of the major repeat 135
Tandem repeat comparison 136
Tandem repetition of *X. laevis* 138
Concerted Evolution 141
Mechanisms of Concerted Evolution 147
Molecular Drive 152

APPENDIX 156

REFERENCES 186

LIST OF RESULTS FIGURES

Fig.	Page	
CHAPTER 3		
R1	73	Rudimentary restriction site maps of major cluster clones.
R2	78	Detailed map of clone XbHW302 with sequence orientation.
R3	83	Spacer sequences comparison between XbHW302 and Xlh3 indicating non-repetitive elements.
R4	84	Repetitive element comparison between XbHW302 and Xlh3.
R5	85	Microinjection of major cluster clones; SDS PAGE.
R6a	87	Microinjection of major cluster clones; TAU PAGE (short exposure).
R6b	87	Microinjection of major cluster clones; TAU PAGE (long exposure).
CHAPTER 4		
R7a	91	Test for repetitive elements in SH115 (short exposure).
R7b	91	Test for repetitive elements in SH115 (long exposure).
R8	92	Southern analysis of *Xenopus* DNAs using SH115.
R9	93	Restriction site map of clone XbHW102.
R10	94	Restriction site maps of clones XbHW9 and XbHW11.
R11	97	Southern analysis of cloned DNA using an H4 gene probe.
R12	98	Southern analysis using partially EcoR1 digested genomic DNA.
R13a	98	The expected product sizes of an EcoR1 digestion.
R13b	98	The sizes and frequencies of an EcoR1 products.
R14	102	A test for unique restriction sites.
R15	102	Southern analysis of Sph1 digested DNA.
CHAPTER 5		
R16	104	Rudimentary restriction site maps of minor cluster clones.
R17	105	Restriction site map of clone XbHW61.
R18	109	H4 and H1 gene silent site divergences within *Xenopus*.
R19	111	Matrix homology plot of XbHW61 against Xlh1 sequences.
R20	112	Microinjection of minor cluster clones; SDS PAGE.
R21a	113	Microinjection of minor cluster clones; TAU PAGE (short exposure).
R21b	113	Microinjection of minor cluster clones; TAU PAGE (long exposure).
R22	114	Southern analysis using an H1 gene probe.
R23	115	Southern analysis of genomic DNA using PH270.
R24	116	Southern analysis of *Xenopus* DNAs using PH270.

LIST OF APPENDIX FIGURES

Fig.	Page	
A1	156	Major cluster clone XBHW302 nucleotide sequence.
A2	158	Translation of XbHW302 gene sequences.
A3	161	XbHW302 H1 amino acid and gene sequence comparisons.
A4	167	Non-repetitive spacer sequence comparisons between XbHW302 and X1h1.
A5	174	XbHW61 genic region nucleotide sequence.
A6	175	Translation of XbHW61 gene sequences.
A7	177	XbHW61 H1 amino acid and gene sequence comparisons.
A8	183	H4 gene comparisons between *X. laevis* and *X. borealis*.

LIST OF TABLES

Fig.	Page	
TI	34	Histone gene numbers in various species.
TII	78	Comparisons of *Xenopus* H1 proteins and genes with *X. borealis* clone XbHW302 H1 gene and predicted translation product.
TIII	137	Comparison of repeat lengths and copy numbers from animals known to display tandemly repeated histone genes.

ACKNOWLEDGEMENTS
I would like to thank Hugh Woodland for his help as supervisor, and Phil Turner for his collaboration, help on the techniques, and allowing me to use his sequence. I am grateful to Clive Wilson for allowing me to screen his library. Thanks also to Bob Old and Mark Boardman for much advice over the last seven years. Many others have also helped in various ways, not least by helping me balance work with relaxation, and seriousness with humour.

DEDICATION
I would like to dedicate this work to my old friend Luke Whiting, and to the people of Langballigholz, a hamlet on the coast of the Baltic.

DECLARATION

The conclusions reached in this thesis are my own, and are based in part on experiments performed in collaboration with others. The preparation of the genomic library, initial screening and isolation of the first clones was performed in association with Dr. P.C. Turner. Likewise the fine mapping of clone XbHW302 and the sequencing of parts of this clone. About one third of the sequence of XbHW302 was determined by Dr. Turner alone, but all comparisons and analyses are my own work. The microinjection studies were performed in association with Prof. H.R. Woodland, with help in some case from Ms. J.E.M. Ballantine.

ABBREVIATIONS

AMPS	Ammonium persulphate
bp	base pairs
BCIG	5-Bromo-4-Chloro-3-Indolyl beta-D galactopyranoside
C	degrees Celsius
CIAP	Calf intestinal alkaline phosphatase
dATP	deoxyadenosinetriphosphate
dCTP	deoxycytosinetriphosphate
ddATP	dideoxyadenosinetriphosphate
ddCTP	dideoxycytosinetriphosphate
ddGTP	dideoxyguanosinetriphosphate
dTTP	deoxythymidinetriphosphate
dGTP	deoxyguanosinetriphosphate
dNTP	deoxyribonucleosidetriphosphate
DTT	dithiothreitol
dTTP	deoxythymidinetriphosphate
DNA	deoxyribonucleic acid
DMF	dimethylformamide
E. coli	*Escherichia coli*
EDTA	Sodium ethylenediaminetetraacetate
g	gram
IPTG	Isopropyl-B, D-thiogalactopyranoside
Kb	Kilobase
mA	milliAmp
Mes	2-N[Morpholino]ethanesulphonic acid]
mg	milligram
mins	minutes
ml	millilitre
mRNA	messenger RNA
Ms222	3-Aminobenzoic Acid Ethyl Ester
ng	nanogram
nl	nanolitre
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PVP	Polyvinylpyrrolidone
rATP	Adenosine triphosphate
RF	replicative form
RNA	ribonucleic acid
SDS	Sodium Dodecyl Sulphate
ss	single stranded
TAU	Triton/Acid/Urea
TEMED	N,N,N',N'-Tetramethylethylenediamine
Tris	Tris (hydroxymethyl) aminomethane, [2-Amino-2-(hydroxymethyl)propane-1,3-diol]
TTP	thymidinetriphosphate
µg	microgram
µl	microlitre

SUMMARY
A previous study showed that unlike *X. laevis*, *X. borealis* contains a predominant or major histone gene cluster, which contains 70% of the 80-90 copies of the H4 genes in the genome (Turner & Woodland 1983 <Nucl.Acids Res.11 978>). This prompted more detailed analysis of histone gene arrangements in *X. borealis*, which is the subject of this thesis. Clones containing histone genes were isolated from a library prepared for this purpose. Analysis of these clones indicated one class containing the major cluster, and a second, minor class which appeared to be cloned at high frequency. Representative members of each class were characterised in detail. Major cluster clone XbHW302 was restriction site mapped, and certain regions sequenced. Microinjection of the major cluster clones in to *Xenopus* oocytes, confirmed that these genes were functional. Attention was then turned to the chromosomal organisation of the major cluster. A 'chromosome walk' experiment allowed the isolation of clones indicating a tandem of arrangement of clusters on the chromosome. This allowed the complete 'repeat' to be mapped in detail. Further studies confirmed that the majority of clusters are tandemly repeated in the genome. The minor cluster clones were analysed in a fashion similar to the major class. Restriction site mapping and DNA sequencing of clone XbHW61 allowed the location polarity, and identity of an H3, H4, and an H1 gene to be determined. Microinjection analysis again showed the genes were functional. Extensive DNA sequence comparisons between various *Xenopus* histone gene clusters were undertaken. The picture of histone gene cluster structure in *X. borealis* that emerges from these studies, is compared at the nucleotide, gene, and chromosomal levels with *X. laevis*. A discussion of how the differences between such closely related species could have arisen is also included.

CHAPTER 1. INTRODUCTION

1.A GENERAL INTRODUCTION

This project arose out of the finding that the histone genes organisation of *Xenopus borealis* was markedly different from that of *X. laevis* (Turner & Woodland 1983). The aim of this study was to characterise in detail, the histone genes of *X. borealis*, so that the significance of, and the mechanisms that could account for these differences might be understood.

The purpose of this introduction is to provide the background information so that this aim can be seen in its wider context. To achieve this the introduction is split into four parts. Part A introduces the frogs, the histones and the histone genes. The bulk of the introduction consists of Section B; a review of the histone gene arrangements in a wide range of species from yeast to man. Section C contains a comparison of the gene numbers of different species. Sequence analysis has yielded information on the fine structure of histone genes, which is discussed in Section D.

1.A.1 INTRODUCING THE FROGS

X. laevis, the South-African clawed toad, and the slightly smaller *X. borealis*, the Kenyan clawed toad, are sufficiently related to be able to form viable hybrids, but these hybrids are sterile. *X. tropicalis* is a more distantly

related relative, and is thought to be more similar to the common ancestor of *X. laevis* and *X. borealis*, than any other living frog. Immunological studies, based on blood albumins have indicated that *X. laevis* and *X. borealis* shared a common ancestor some 8-10 Myr ago (Bisbee *et al* 1977). More recently, comparative sequence analysis of adult globin cDNAs has suggested a figure of 15-20 Myr ago (Knochel *et al* 1986). The authors of the latter study argue that as an antigenic site is only a fraction of a molecule, a comparison of the entire coding sequence is likely to be more accurate.

1.A.2 INTRODUCING THE HISTONES

Histones have been studied for over a century (Kossel 1884). They are a class of small basic proteins that associate with each other and with DNA to form the nucleosome, the fundamental unit of chromatin. Together with other nucleoproteins, nucleosomes form into higher orders of chromatin structure. Histones are found in every eukaryote, and some histone-like proteins have been identified in a variety of prokaryotes (Lathe 1980, Hubscher *et al* 1980, Delange *et al* 1981). The four nucleosome core histones, H2A, H2B, H3, and H4, occur in equimolar quantities, but H1, the fifth histone type, occurs in half the molar equivalent of the core histones in higher eukaryotes. In yeast a homologue of H1 has not been identified.

Introduction

The nucleosomal core contains an H3-H4 tetramer which is associated with two molecules each of H2A and H2B (Kornberg 1977). Around this is wrapped some 146 bps of nuclear DNA in 1.7 turns. The complete nucleosome particle also contains a single H1 histone (McGhee & Felsenfeld 1980). It appears that the spacing of nucleosomes along nuclear DNA varies between species and between cell types within a species. The structures of histones, especially of H3 and H4 are highly conserved between the animal and plant kingdoms (Panyion et al 1971). The H4 histone of pea plants differs from that of calf thymus by only two conservative replacements (Delange et al 1969). However, the H1, H2A, and H2B histones display less conserved peptide sequences. In chicken, a cDNA clone coding for a highly diverged H2A (H2A.F) has been identified (Harvey & Wells 1984). In terms of function, it seems likely that the nucleosome must have a highly conserved core for maintaining compaction of the DNA. Within any histone type the available positions for change, which can be accommodated without loss of function, are indicated by the heterogeneity that is found within and between species. The H1 histone shows greatest divergence, but nevertheless contains a conserved central hydrophobic region around amino acids 93 - 110. The H1 histone is also the largest protein.

Histone proteins can undergo several post-synthetic modifications, namely acetylation, phosphorylation, methylation, ADP-ribosylation, and in the case of histone H2A, covalent attachment to the protein

ubiquitin, to produce protein A24. The exact role of these modifications is not well understood, but they are thought to play major roles in various nuclear functions.

1.A.3 INTRODUCING THE HISTONE GENES.

Histone genes were amongst the first genes cloned (Kedes 1975a). Furthermore, even before cloning, the genes had been studied after purification of genomic DNA by caesium chloride gradient centrifugation (Birnstiel *et al* 1974). Today they rank among the most studied gene types.

Histone genes fall into two categories. The first class tend to be cell cycle regulated, although in early amphibian embryogenesis this is not the case. They typically form multigene families, have no introns, and display a hyphenated dyad symmetry sequence at their 3' end which is involved in an unusual, polyadenylation-independent, mRNA processing mechanism (reviewed Birnstiel *et al* 1985). The second class are the so called 'replacement' histone genes. These genes are expressed in a low, constitutive, manner, which allows their accumulation in non-dividing cells. Their mRNAs generally are polyadenylated, and show regulation features similar to ordinary protein coding genes (Trainor & Engel 1989). The best studied example of this latter class is the chicken H5 gene (Krieg *et al* 1983), although the H2A.F falls into this category as well. The *Xenopus* histone genes described in this thesis fall into the former, cell-cycle regulated category.

Introduction

The involvement of histones in gene regulation is controversial, but interestingly histone variants may be synthesized at different times during embryonic development, or during the cell cycle. Histone gene families may be regulated by complex controls ensuring that proper stoichiometric relationships between histones and DNA are maintained, and that different variants are expressed appropriately. These controls have been the subject of intensive research that as well as focusing on the metabolism of histone gene products, has also concentrated on the structure and evolution of histone genes themselves.

1.B.1 INTRODUCTION TO HISTONE GENE STRUCTURE

An understanding of histone gene fine structure, and the arrangement of histone genes on the chromosome is a starting point toward comprehending the mechanisms controlling expression of these genes. Molecular cloning has allowed the isolation of many histone gene clones from a wide range of both lower and higher eukaryotes. Analysis of these clones has led to a detailed understanding of histone gene organisation within many individual species (reviewed Wentschel & Birnstiel 1981, Maxson *et al* 1983c). Histone genes can occur singly, or clustered together, and may occur in tandemly repeated clusters containing one gene for each of the five histone types.

Together with copy number, the various arrangements differ so markedly between closely related species, that evolutionary trends in histone gene organisation are obscure (Maxson *et al* 1983b). This particular point is addressed in this thesis. Sequence divergence between variant genes appears almost as diverse. Fine analysis of the histone gene structure of variants known to be expressed at different developmental stages or in different tissues, is beginning to help elucidate the mechanisms of control. (eg Trainor & Engel 1989).

It is to be hoped that as studies on histone gene structure and expression proceed hand in hand, the mechanisms of regulation, as well as evolution will continue be elucidated.

1.B.2 *STYLONYCHIA MYTILIS* AND *SACCHAROMYCES CEREVISIAE*.

Two Protists that have been studied are the protozoan ciliate *Stylonychia mytilis*, and the yeast *Saccharomyces cerevisiae*. The macronuclear DNA of *Stylonychia mytilis* exists as discrete fragments of 0.45 to 25 Kb (Lipps & Steinbruck 1978). Southern hybridisation using histone gene probes revealed that each probe reacted with a single sized fragment of DNA. Therefore, the modestly repeated histone genes do not appear to be clustered (Elsevier *et al* 1978).

The first yeast histone genes to be isolated were the H2A and H2B (Hereford *et al* 1979). Subsequently, clones containing H3 and H4 genes were isolated independently

(Smith & Murray 1983). Unlike any other eukaryote investigated, no H1 protein, nor gene, has yet been isolated despite considerable effort (Sommer 1978, Mardian & Isenberg 1978, Brandt *et al* 1980). After such extensive investigations, it is likely that none exists.

Analysis of the clones initially isolated, showed that the two clones carrying H2A and H2B genes differed in restriction site locations, but both contained H2A and H2B genes, closely linked to each other, and certain non-histone protein genes. Analysis of H3 and H4 gene clones showed these two genes were closely linked, and like H2A and H2B, divergently transcribed, but unlinked to H2A and H2B genes. Again two different H3-H4 gene pairs have been identified. The distance between each member within the four pairs of genes is 600-800 bp. Restriction site and DNA sequence analysis, suggests that the pairs of H2A-H2B genes and H3-H4 genes are non-allelic. The H3-H4 gene pairs predictably show stronger sequence homology than the H2A-H2B gene pairs. These clusters together represent all of the eight *S. cerevisiae* histone genes, which exist in the haploid genome. Evidence for this came from genetic mapping, genome restriction site mapping, and Southern blots.

1.B.2a. *CAENORHABDITIS ELEGANS*

Analysis of cloned and genomic DNA from the nematode *C. elegans* indicates approximately eleven dispersed histone gene clusters (Roberts *et al* 1987). Studies on four cloned

clusters, reveal the histone genes vary in number and polarity (Roberts *et al* 1989), ruling out a tandem repeat arrangement.

1.B.3. SEA URCHINS

Sea urchin histone gene organisation and expression has been extensively studied in many laboratories, partly due to the early cloning of these genes (Kedes 1975a).

Histone gene expression in sea urchins occurs in three developmental stages, from three different sets of histone genes (reviewed Kedes 1979). The early genes code for the predominant histones of egg and early embryo. A second set of H1, H2A, and H2B histone proteins are also found in the egg. These cleavage-stage variants are synthesised by the embryo for only a few hours following fertilisation. They are products of a still uncharacterised set of histone genes. Later embryos synthesise a third set of histones, which are products of transcription from a third set of histone genes; the late histone genes.

EARLY GENES

Recombinant histone gene clones were initially isolated from a genomic sea urchin DNA library by the use of labelled mRNA probes (Kedes 1975a). Since then clones from five sea urchin species have been isolated, namely *Psammechinus milliaris* and *P. lividus*, *Lytechinus pictus*, *Echinus esculentus*, and *Strongylocentrotus purpuratus*. Analysis of these clones has

yielded a wealth of information. Firstly, sea urchin early histone genes appear clustered together into a structure known as a repeat. These repeats take their name from the fact that there are many copies in each genome. *P. milliaris* has been shown to have one repeat of 6.3 Kb reiterated 300-600 fold, but another repeat of 6.7 Kb, at a much lower frequency (5-10 copies) (Gross *et al* 1976, Birnstiel *et al* 1978, Busslinger *et al* 1980). *L. pictus* is believed to carry some several hundred copies of a 7.2 Kb repeat (Cohn & Kedes 1979a, 1979b), and *S. purpuratus* has an estimated 300 copies of a 6.5 Kb repeat (Kedes 1979, Overton & Weinberg 1978).

Certain features have been found to be common to all early sea urchin histone gene repeats studied. Each repeat has been shown to contain one copy of each histone gene, each of which is transcribed off the same strand, ie. each gene displays the same relative polarity. In every case the gene order is H4>-H2B>-H3>-H2A>-H1>. Although the genes are transcribed off the same strand, they are not transcribed to produce a single RNA species, but rather form five transcripts. The repeats are arranged in a tandem fashion, forming a battery of serially reiterated clusters. This was consistent with the belief that the histone genes occurred at a very few loci, as histone gene DNA was already known to produce a satellite when separated on a caesium chloride density gradient (Birnstiel *et al* 1974).

Introduction

As it is now firmly established that several isotypes of histone proteins H1, H2A, and H2B are expressed in early embryogenesis (Cohen *et al* 1975, 1978, Brandt *et al* 1979), the highly reiterated histone gene batteries must be envisaged as containing a certain number of gene copies for the subtypes of the different histones. Work on *S. purpuratus* has shown that at a sequence level each repeat is not identical, and that the spacer regions in particular, display considerable sequence heterogeneity (Overton & Weinberg 1978).

SEA URCHIN LATE GENES

Electrophoretic separation has allowed comparison of histones from different developmental stages. This revealed different H1 histone isotypes to be present between blastula and gastrula stages (Seale & Aronson 1973, Ruderman & Gross 1974). A similar result was also later found for H2A and H2B histones (Cohen *et al* 1975, Newrock *et al* 1978). Furthermore these changes could be directly correlated with changes in both size and sequence of histone mRNAs from different stages (Grunstein *et al* 1976, 1981, Grunstein 1978, Kunkel & Weinberg 1978, Childs *et al* 1979, Meiter *et al* 1979). Analysis of the genes coding for these late histone proteins was retarded partly due to their relatively minor abundance (Kunkel & Weinberg 1978, Childs *et al* 1979).

Clones carrying late genes have been isolated and analysed both from *S. purpuratus* (Maxson *et al* 1983a) and *L. pictus* (Childs *et al* 1982).

Three *L. pictus* clones were isolated by null restriction cloning (see section 1.B.12) to avoid cloning early repeats. These clones each contained an H3-H4 gene pair. Restriction site mapping showed that these represented three different gene pairs. They were confirmed as 'late' genes by both hybrid release translation, and RNA melting curve analysis.

Originally there were thought to be 20-50 fold fewer late genes than early genes (Kunkel & Weinberg 1978, Childs *et al* 1979), but Southern blot analysis using late histone gene hybridisation probes on genomic DNA led to an estimate much lower than this, namely in the order of 10 or less copies per haploid genome. This analysis also indicated considerable sequence heterogeneity in the flanking regions between the five individuals tested. The sequence of one clone revealed the following three points of interest. Firstly, the genes were divergently transcribed, in a fashion similar to the histone genes in yeast. Secondly, the predicted amino acid sequence of both H3 and H4 genes was identical to the early histone genes. Thirdly the non-coding 3' trailers and 5' leaders showed little or no homology to those of the early genes, except for certain expected regions (see sections 1.C.2 and 1.C.3). These genes were not

found in close proximity to H1, H2A, or H2B histone genes, and the clones also showed differing intergenic spacer regions between the different copies.

The late histone genes were isolated from *S. purpuratus* utilising a different approach, and resulted in the isolation of late H2B and H4 genes. Analysis of these clones revealed divergent polarities, and copy number of 5-10. Thus the sea urchin later genes are markedly different from the early genes.

1.B.4 *DROSOPHILA*

A Col E1 recombinant plasmid library containing *D. melanogaster* genomic DNA was prepared and was screened using a sea urchin mRNA hybridisation probe (Lifton *et al* 1978). From this the plasmid cDM500 was isolated. Initial analysis of this clone revealed it contained a sequence which was tandemly repeated.

The presence of histone genes within this repeat, and the determination of types present was performed as follows. *Drosophila* poly(A) minus RNAs were extracted from embryos and tissue culture cells, and were purified by hybridisation to bound cDM500 DNA, followed by gel electrophoresis. These separate purified mRNAs were then used as hybridisation probes on Southern blots of restriction digested cDM500 DNA. This revealed that all the genes for these mRNAs were present on a single repeat cluster. Identification of each histone type was obtained by partial DNA sequencing of the regions homologous to the RNAs.

The polarity of each gene was determined by annealing separate histone mRNAs to separated strands of a hybrid lambda phage containing the entire 4.8 Kb cluster. This supported the partial DNA sequence analysis. The polarity and order was thus established as <H1--H4>--<H2A--H2B>--<H1. Studies on the reassociation kinetics of the cloned repeat in the presence of a vast excess of genomic DNA produced an estimate of 110 copies

(Lifton et al 1978). Further analysis of genomic clones revealed that there were only two types of repeat, differing only by the insertion of a 240 bp fragment in the downstream spacer region of the H3 gene (Goldberg 1979). Those carrying the extra insert have been classified as long (L) repeats, those without, short (S).

The ratio of L to S types of the repeat has been investigated by blot hybridisation, which indicated a ratio of 3:1 respectively (Lifton et al 1978). Further analysis of the organisation of the repeats utilised a EcoR1 restriction site present in the long repeats, but absent in the short units, and a BamH1 site present in both. Comparisons of restriction digested genomic DNA followed by two-dimensional electrophoresis blot hybridisation and densitometry allowed analysis of the relative positions of L and S units. In conclusion, there appear to be, only two LSL arrangements, one LSSL, four to five of the type LSSSSL or LSSSSSL, plus a further ten arrays containing the remaining L units; an average of six per array. This analysis indicates the existence of several non histone insertions of arbitrary length, but which must contain EcoR1 and Bam H1 sites. Any functional significance of these arrangements, and of the number and positions of the insertions, is unknown.

Thus histone gene arrangements in *Drosophila* appear similar to those of the sea urchin early histone genes. Interestingly, restriction site analysis has not detected any other sequence heterogeneity corresponding to the 'late' type histone genes of the sea urchin.

In situ studies have established that the histone genes extend from 39D2-3 to 39E1-2 on the left arm of chromosome 2 in *D. melanogaster* (Pardue *et al* 1977). A nick translated *D. melanogaster* histone probe was also tested for *in situ* hybridisation to *D. miranda*, and *D. pseudoobscura*, a close relative of *D. miranda* (Steinemann 1982). The left arm of chromosome 2 in *D. melanogaster* is expected, based on all other homology correlations, to correspond to chromosome 4 in *D. miranda* (and *D. pseudoobscura*). However the experiment revealed hybridisation to chromosome 2 in both *D. miranda* and *D. pseudoobscura*. One possible explanation is the translocation of the histone gene cluster in a common ancestor of *D. miranda* and *D. pseudoobscura*. Taxonomically *D. miranda* is very closely related to *D. pseudoobscura*, and both are in the *obscura* group, whereas *D. melanogaster*, is in the *melanogaster* group. Hybrids between these different groups cannot be formed. Interestingly, the chromosomal location of the 5S RNA genes, which belong to the same class of moderately reiterated, tandemly repetitive DNA as the histone genes, was not on the expected analogous chromosome. Probes from four other DNA fragments of *D. melanogaster*, which originated from low or single copy sequences, all

showed homology to the expected analogous chromosomal locations in both *D. miranda* and *D. pseudoobscura*. This might indicate a greater cytogenetic mobility of tandemly repeated genes.

One possible explanation, is that these changes may have arisen due to unequal cross-over events. This explanation is favoured because tandem repetition is thought to increase the frequency of cytogenetic movement, by the mechanism described below. Unlike low copy elements, tandemly repeated sequences invariably display regions of homology along the chromosome. Mispairing between homologous regions at different positions within the arrays of two chromosomes pairing at meiosis and mitosis, could then result in chromosomal misalignment. Unequal cross-over events between such mispaired chromosomes would then produce exchange of repeated sequences.

1.B.5 *ARTEMIA*

Artemia salina, the brine shrimp, acts as a useful comparison with *Drosophila*, as it also belongs to the same phylum in the larger phylogenetic group of the protostomes, one of the two major groups of animal phyla that diverged more than 600 million years ago.

Again the histone gene organisation of this species has been characterised by analysis of clones obtained from a genomic library, in this case using the *Drosophila* probe cDM500 (Bagshaw *et al* 1984). Restriction

site mapping of four clones displayed two histone gene arrangements. Three clones displayed the same restriction site pattern. Southern analysis of restricted genomic DNA, revealed that not only were nearly all the *Artemia* histone genes on this cluster type, but it was also shown that they were tandemly repeated. This was based on the observation that BamH1 only cut the cloned DNA at one site per cluster, while HindIII, cut at several closely spaced sites. A Southern blot of genomic DNA digested with each enzyme separately was probed with a plasmid subclone carrying all the histone genes. BamH1 digestion produced a single band of 8.5 Kb. The HindIII digest produced distinct fragments of sizes also consistent with the cloned cluster map. However, as the largest 7.8 Kb HindIII fragment overlaps the single BamH1 site, it is clear that the cluster must be tandemly repeated, with a repeat length is 8.5 Kb. The gene order and polarity has yet to be determined. Analysis of the other clone, a rare variant, revealed a gene order of H2A-H4-H3-H2B, which is different to both *Drosophila* and the sea urchin. The use of a *Drosophila* H1 histone gene probe did not allow the identification of *Artemia* H1 histone genes, so the occurrence of *Artemia* H1 histone genes has yet to be investigated.

1.B.6 RAINBOW TROUT

The rainbow trout, *Salmo gairdnarii*, is the most primitive vertebrate for which the histone gene arrangements have been studied. A *Xenopus* H4 histone gene subclone was used to screen a total genomic DNA library in Lambda Charon 4A. Analysis of twelve clones isolated from this library, showed that each contained a common 10.2 Kb EcoR1 fragment, which contained one copy of each histone gene type. The order and polarity was determined as H4>--H2B>--H1>--H2A>--H3> (Conner *et al* 1984). However one clone displayed an apparent 400bp deletion in the region corresponding to the H1 gene in one of the twelve clones. The location of the genes was determined by using various histone gene probes on Southern blots, carrying different digests of the cloned repeat. The direction of transcription was determined by Maxam and Gilbert sequencing. Southern blots of restricted genomic DNA revealed that the majority of the trout histone genes occur on this 10.2 Kb cluster type.

Copy number was determined by dotting known quantities of genomic and cloned DNA on to nitrocellulose filters. These were then probed with a histone gene hybridisation probe. Scintillation counting was used to determine the amount of hybridisation. The value per sperm nucleus was determined to be 145 copies. As rainbow trout are thought to be tetraploid (Ohno 1970) this would indicate about 73 histone genes per haploid genome.

There is no evidence to suggest that these genes are tandemly arranged: inspection of the restriction site maps of the clones indicate no common restriction sites outside the 10.2 Kb. unit in any of the twelve clones which extend up to 8 Kb. on either side of the 10.2 Kb. repeat.

1.B.7 NEWT

Notophthalmus viridescens, the American spotted newt, is the only urodele for which the histone gene organisation is well characterised. Clones were obtained on the basis of hybridisation to a previously identified newt H4 pseudogene (Stephenson *et al* 1981a). A 9.0 Kb DNA segment was found to be common to several clones and so underwent further analysis. It was shown to contain one copy of each of the five histone genes. The positions of the H2A, H2B, H3, and H4 genes were determined by hybridisation with cloned sea urchin histone genes, and were further confirmed by partial sequence analysis, which also indicated the direction of transcription. The gene order is H1>--H3>--<H2B--H2A>--H4>, with each gene except H2B being transcribed in the same direction (Stephenson *et al* 1981a).

Analysis of Southern blots of restriction digested genomic DNA, using an H4 histone gene probe revealed an apparent copy number of 600-800 per haploid genome, and that the majority of these occurred on fragments of the sizes predicted from the spacing of the restriction sites in the cloned cluster (Stephenson *et al* 1981b, Stephenson 1984).

Minor bands were interpreted as occurring from cases of heterogeneity in the position or occurrence of a single site. In conclusion, the newt histone genes occur on a highly reiterated cluster, which shows a constant gene structure, and an almost identical restriction enzyme site pattern. The occurrence of a site, not present in the cluster allowed demonstration that the distance between the clusters exceeded 25 Kb. The spacer DNA between them seems to contain highly repetitive satellite DNA. *In situ* studies using histone probes reveal only two chromosomal loci contain histone genes. These sites also strongly hybridise to highly repetitive satellite DNA probes (Diaz *et al* 1981).

The arrangement described here appears unique to the newt. All other species either display tandem repetition to form a battery of 'histone DNA' with relatively short regions between repeats, or a more dispersed arrangement, with histone genes at various loci.

1.B.8 *XENOPUS*

Histone gene arrangements in three species of *Xenopus*, *X. tropicalis*, *X. laevis*, and *X. borealis* have been studied. *X. tropicalis* differs from the other two species in that it displays 20 chromosomes [1n=10] while *X. laevis*, and *X. borealis* display 36 [2n=18](Tymowa & Fischberg 1973). It is

widely accepted that this is due to a genome duplication event in the lineage that gave rise to *X. laevis* and *X. borealis* (Bisbee *et al* 1977).

The least studied of these species is *X. tropicalis*, from which only two clones have been isolated (Ruberti *et al* 1982). The gene order of the cluster present on these clones differs. Blotting of genomic DNA shows that a number of DNA fragments react with radiolabelled histone DNA probes, indicating several cluster types.

X. laevis has been most extensively studied, by various groups (Zernik *et al* 1980, Ruberti *et al* 1982, Old *et al* 1982, Destree *et al* 1984), but certain controversies remain. The original estimate of reiteration frequency (Jacob *et al* 1976), based on DNA/RNA hybridisation kinetics, indicated a value of between 20 and 50 copies per haploid genome. This was disputed by Sommerville (1979), who produced estimates of 50-86. Van Dongen *et al* (1981) followed, returning to a lower value of 45-50 copies, which again was disputed by Turner and Woodland (1983), who calculated a value of 80-90 at a defined hybridisation stringency. A later paper (Perry *et al* 1985) again proposes a lower value of 45-55, but under more stringent hybridisation conditions.

Views on histone gene organisation have had a similar history. Van Dongen *et al* (1981), originally showed that two genomic EcoR1 restriction fragments, 8.9 and 5.1 Kb in length reacted strongly with histone DNA probes. Several

DNA fragments showing weaker hybridisation signals were also observed. These results led to the proposal that most *X. laevis* histone genes are organised in a cluster, 14 Kb in length, that is cleaved by EcoR1 into two fragments of 8.9 and 5.1 Kb in length, and has a gene order of H4-H3-H2A-H2B. Turner and Woodland (1983) advanced several arguments supporting an alternative arrangement. First, they pointed out that the postulated 14 Kb EcoR1 fragment described had not been found among the 21 *X. laevis* clones examined at that date (Old *et al* 1982, Ruberti *et al* 1982, Turner & Woodland 1982), but rather there was considerable heterogeneity in the cloned EcoR1 fragment sizes containing histone coding sequence. The most common EcoR1 fragment in existing clones is 8.6 Kb in length and contains genes for all five histones (Old *et al* 1982). This fragment could be a simple size variant of the 8.9 Kb fragment, as it displays common features. The most common gene order observed in the clones is H1-H2B-H2A-H4-H3, or circular permutations of this (Old *et al* 1982). The original order, which did not include an H1 gene, has only been observed once. Lastly, the relative amounts of 8.9 and 5.1 Kb genomic fragments vary extensively between individuals, supporting the idea that they are unlinked, and therefore could not combine to form a 14 Kb repeat unit (Turner & Woodland 1983).

Turner and Woodland therefore proposed an updated arrangement in which there are two major cluster types, each with a different gene order H1-H2B-H2A-H4-H3, and

H4-H3-H2A-H2B, and each containing less than half of the histone gene complement in the genome. This pattern was exhibited by seven clones isolated by Old *et al* (1982). Several of the more recently isolated *X. laevis* histone gene clones also display these arrangements (Destree *et al* 1984, Perry *et al* 1985), but the distances between genes, and the transcription polarities vary widely between clones of the same gene order. Thus two such clones can easily appear very dissimilar, and unrelated.

The amount of homogeneity within *X. laevis* has been studied. Turner and Woodland (1983) claim different observations for similar restriction enzyme digestions from Van Dongen *et al* (1981), and conclude that different *X. laevis* individuals seem to be quite variable in their histone gene cluster structure, and support this with reference to the variability of gene cluster structure found in the characterised clones. This is in accord with the early findings of Old *et al* (1982). Old *et al* analysed seven genomic clones, which were found to display heterogeneity in restriction site positions of both coding and non-coding regions, and concluded that *X. laevis* is highly heterogeneous, and does not display histone genes organised into identical repeating units. This is in contrast to the work of Van Dongen *et al* (1981), who proposed about 60% of the histone genes in *X. laevis* occur on a cluster of H4-H3-H2A-H2B. Recent work (Perry *et al* 1985) supports the observations of Van Dongen *et al*, and concludes that there

is a major cluster carrying the majority of the histone genes, in conjunction with most, if not all the H1A genes. Interestingly the results of Perry *et al* are based on analyses using DNA pooled from more than 24 individuals. This has a highly significant effect. The polymorphism displayed between individuals is likely to be minimised by this process. For example, if each of the twenty four individual frogs had a single cluster common to all, and a second cluster, which was unique to each individual, Southern blot analysis of pooled DNA would tend to indicate only the presence of the common cluster. Clearly relative to each individual, the multiplicity of over-representation of the common cluster is equal to the total number of pooled samples displaying the common cluster.

The idea that different H1 genes are present on different cluster types has also gained recent support. It was first proposed in 1980 (Zernik *et al* 1980), with the isolation of three clones of two gene orders. Hybrid release translation revealed clusters of one gene order carried H1B genes, while the other carried the H1A type. This finding, and that the cluster orders for the H1A type and H1B type are the two proposed by Turner and Woodland has been confirmed (Perry *et al* 1985, Destree *et al* 1984).

1.B.8b *Xenopus borealis*

Southern analysis of genomic DNA, using histone gene probes showed the majority of *X. borealis* histone genes occur on a single cluster type (Turner & Woodland 1983). Hybridisation studies revealed that *X. borealis* and *X. laevis* had roughly similar numbers of H4 genes (80-90, see above), and that in *X. borealis* 70% of the these were in the major clusters. The restriction site pattern of this cluster was sufficiently conserved to undertake its crude genomic mapping. This revealed a novel cluster structure. Clearly the histone gene organisation in *X. borealis*, contrasts strongly with that in *X. laevis*. This observation prompted the research work presented in this thesis. The aim, therefore was to characterise the histone genes of *Xenopus borealis* fully , and to try to account for the observed differences with *X. laevis*.

1.B.8c HISTONE GENE EXPRESSION IN *XENOPUS*

What is the functional significance of the diversity in *Xenopus* histone gene cluster structure? In sea urchin there is a link between structure and function, in the form of different temporal expression from different types of clusters (Section 1.D). Woodland *et al* (1984) asked the question whether *Xenopus* displays differential spatial or temporal expression of histone genes, by sequencing the leaders of various H4 mRNAs. This was done as follows. Radiolabelled primers complimentary to the front coding

portion of an H4 gene were hybridised at different temperatures to mRNAs from different tissues of several individuals. After reverse transcription, the products were separated by gel electrophoresis, and detected by autoradiography. The different products were then cut out of the gels, and sequenced. In *X. borealis* a single predominant class was found in all tissues studied (ovary, tadpole, lung primary culture, heart primary culture). The members of this class all displayed the same sequence, but the products appeared to vary in length. When ovaries of different individuals were studied, a single base change was found. At the lower hybridisation temperature, a second band was also seen in all tissues, and in every individual. The sequence of this band showed clear differences from the major band. Again two sequence types were found. Clearly, there appears to be neither a developmental switch, as seen in sea urchins, nor tissues specific expression. A second, subsidiary aim of this project, was to determine whether the predominant mRNA class was coded for by the major gene cluster, and to characterise the genes of the minor mRNA class, found at the lower hybridisation temperature.

So to conclude, *Xenopus* shows both similarities and differences with the primitive invertebrates described in the previous section, in that this genus displays heterogeneous arrangements of histone genes (*X. laevis*, and

X. borealis), as well as repeated clusters of a conserved type (*X. borealis*).

1.B.9 BIRDS

The only published data relating to the histone gene structure and organisation of birds are those for the chicken *Gallus domesticus*. About 20 recombinant histone gene clones have been examined (Engel & Dodgson 1981, Engel *et al* 1982, Harvey *et al* 1982, Krieg *et al* 1982, Sugarman *et al* 1983). These are estimated to comprise 70% of the genomic complement of 10 genes each for the five histone types (Sugarman *et al* 1983). These clones average approximately 15 Kb in length, and carry from one to six histone genes, at various intergenic distances. They display a wide range of gene orders, and show differing directions of transcription. This eliminates the idea of a repeating gene order, although it has been clearly demonstrated that a few clones overlap with other recombinants. The results of efforts to 'walk' down the chromosome, led to the discovery of 10 histone genes on 36 Kb of contiguous chromosomal DNA (Harvey & Wells 1984). These are clustered in three groups of 5, 3, and 2 histone genes separated by long intergenic spacers, sometimes longer than 10 Kb. The lack of a regular genomic arrangement is further supported by genomic DNA Southern blot analysis, which shows that a number of genomic DNA fragments of variable size react with radiolabelled histone probes (Engel & Dodgson 1981, Sugarman *et al* 1983).

1.B.10 MAMMALS

The histone genes of mammals appear, like those of the chicken, to have a copy number of about 20 per haploid genome. This figure has been determined for mice using solution hybridisation (Jacob 1976), and is in agreement with blot analyses for both mouse and human. In the mouse, the majority of these genes map to chromosome 13, while the rest map to chromosome 3 (Marzluff & Graves 1984). The structures of several mouse and human histone gene clones have been reported (Seiler-Tuyns & Birnstiel 1981, Heintz *et al* 1981, Sierra *et al* 1982, Sittman *et al* 1981). Mapping and sequence analysis for the human clones, and subsequently those of the mouse as well, showed that some of the genes were linked in the genome. Further analysis of the clones revealed that the genes could vary in order, spacing, and histone type identity. Thus, in neither species are histone genes arranged in a repeated manner. This heterogeneous organisation is confirmed by genomic DNA gel blot experiments. A mouse H4 gene probe reacts with 11 genomic restriction enzyme fragments of different sizes (Seiler-Tuyns & Birnstiel 1981). The apparent lack of any human clone displaying all five classes of histone genes together, given that the average length is 15 Kb, argues against the existence of a repeating unit. Five different gene orders have been identified in the human histone clones. Restriction site analysis of spacer DNA, flanking genes and spanning the intergenic regions shows these

sequences to be highly diverged (Maintz *et al* 1981). The length of the spacer DNA, between histone genes on the various human histone clusters is also variable. Analysis of this spacer using cloned AluI DNA, has indicated the presence of repetitive DNA interspersed between genes (Sierra *et al* 1982). *In situ* studies using histone hybridisation probes have shown these genes to be located at the distal end of chromosome 7 in the region 7q32-36 (Chandler *et al* 1979).

1.B.11 ORPHONS

Orphon genes are defined as dispersed, solitary genetic elements derived from multigene families (Childs *et al* 1981). As orphon genes occur in more than one species, they are introduced separately.

The isolation of a histone orphon arose out of an attempt to isolate the late sea urchin histone genes of *L. pictus* (see section 1.B.3), by the process of null restriction cloning. Null restriction cloning allows the cloning of rare variant elements (eg sea urchin late or orphon histone genes), from a DNA preparation containing sequences to which the hybridisation probe will also bind, and which may be so abundant (eg sea urchin early genes in a total sea urchin DNA preparation) as to reduce the chance of isolation of the desired elements by the usual techniques. This elegant selective exclusion is achieved by the use of a

restriction enzyme which does not allow fragments of a size small enough to be cloned to be formed from the common elements.

In the case of *L. pictus*, BamH1 does not cleave the early repeats, which are tandemly arranged, and so is ideal. This technique was tested by first separating, then Southern blotting the cloned DNA followed by hybridisation to the sea urchin early histone gene probe. This allows the null restricted DNA to be visualised at a high molecular weight.

The distinction between orphon and late genes was achieved by RNA/DNA hybridisation, as cross hybrids between early and late dissociate at a lower temperature. These results indicated that the isolated H3 orphon displayed a nucleotide sequence that was identical or nearly identical to the early H3 genes. R-loop analysis showed that the H3 orphon was at another locus relative to the early genes. Heteroduplex analysis to the cloned early DNA revealed that at least 1.7 Kb of 5' sequence and 1 Kb of 3' sequence was also cloned. This also revealed the presence of an inverted repeat sequence some 130 bp long in the 3' flanking region. Further restriction site, and sequence analysis showed that the isolated orphon gene displayed only 1 % divergence from the early pLpC clone, and therefore had only diverged very recently in evolutionary terms.

Genomic Southern blot experiments revealed that at the 5' and 3' ends, different moderately repetitive DNA occurs. The boundaries of the histone gene insert did not display the terminal repeats associated with transposable elements, indicating some other mechanism of generation. Comparisons of six individuals using genomic southern blots revealed that no two individuals displayed the same pattern for any of the five orphon histone gene types investigated.

The mechanism of generation of orphons remains an enigma. The lack of terminal repeats argues against some method involving an autonomous transposon, or a transposase. There also appears to be no evidence to support transfer via an RNA intermediate (see Maxson *et al* 1983c). Removal of a gene to a different environment, could easily change the regulation pattern of that gene, which may then be under different selective pressures, and may be free to evolve independently. Temporal or special regulation of histone genes may have arisen this way. However there is no evidence that orphon histone genes behave in this way.

1.C HISTONE GENE NUMBERS

It is apparent that organisms with a large genome must have more DNA to package, and so consequently require more histones. This has led investigators to seek to display a correlation between the genome size (C-value) and the number of histone genes.

As mentioned earlier, in the case of *Xenopus* (see section 1.B.8) disagreements have arisen as to the exact number of genes present per haploid genome. This hinders comparisons: clearly wherever there is sequence heterogeneity within the histone genes of an organism, a homologous probe (usually produced from cloned DNA) used at high stringency will produce an under-estimate of the histone gene numbers. This argument has been levelled at several workers (Jacob *et al* 1976, Van Dongen *et al* 1981), by Turner and Woodland (1983).

Even though H4 genes have been chosen in many cases, for the reason that it is the most highly conserved (eg Hilder *et al* 1981, Turner & Woodland 1983), the nucleotide sequence can vary up to 26% between species, due almost entirely to 'wobble' in the third bases (Busslinger *et al* 1982). It has even been suggested that homologies between extreme variants may be so weak as to preclude cloning of a particular variant by using an 'early' gene (Harvey & Wells 1984).

Table I shows the histone gene numbers for some species. There appears no clear, direct correlation between histone gene numbers and the C-value, although it is true that organisms with high C-values have a large number of histone genes. Organisms with a large histone gene complement, on the other hand, do not necessarily have large genomes. Birnstiel *et al* (1979) suggested that the high number of histone genes in the sea urchin is necessary to

HISTONE GENE NUMBERS

SPECIES	COPY NUMBER	REFERENCE
Ambystoma mexicanum (Axolotyl)	2685 +/-349	Hilder *et al* 1981
Necturaus maculosus (Mudpuppy)	1200 - 2000	Sommerville 1979
Notophthalmus viridescens	600 - 800	Stephenson *et al* 1981a
Triturus sp.	350 - 640	Hilder *et al* 1981 Sommerville 1979
Sea Urchin	300 - 500	Kedes & Birnstiel 1971 Kedes *et al* 1975
Drosophila	110	Lifton *et al* 1977 Pardue *et al* 1977
X. laevis	50 - 90	Hilder *et al* 1981 Turner & Woodland 1982
X. borealis	50 - 90	Turner & Woodland 1982
Salmo gairdnarii (Rainbow trout)	73	Conner *et al* 1984
Human	30 - 40	Wilson & Melli 1977
Mouse	10 - 20	Jacob 1976
Chick	10	Crawford *et al* 1979
Yeast	2	Wallis *et al* 1980

TABLE 1 Histone gene numbers in a variety of species.

[Figures for one core histone gene type per haploid genome]

satisfy the requirement of histones during cleavage. Mammals and birds display fewer histone genes, and do not have such high cell division rates during embryogenesis. Either with or without a rapid cleavage stage in early development, different strategies can help explain a lack of correlation between different species. One might expect that within a group, such as the amphibians, the members may display similar developmental patterns, and consequently show the correlation as described. A better correlation than for all species has been noted in a comparison of three Amphibia (Hilder *et al* 1981), but it was concluded that additional factors were required to account for the observed differences.

It is unclear if all the genes are functional; histone pseudo-genes have been isolated (Turner *et al* 1981), but are rare. There are also 'orphon' histone genes, which could be expressed incorrectly, or not at all. Clearly non-functional genes would mask any true correlation between C-value and histone gene number. On the other hand, it has been calculated that *Xenopus* has around the minimum number of genes necessary to allow the accumulation of sufficient histone message and histones during oogenesis to package the large amounts of DNA which are rapidly synthesised during early embryonic cleavage (Woodland 1980).

1.D FINE STRUCTURE OF HISTONE GENES

The previous section describes the histone gene organisation at the cluster level. DNA sequence analysis allows a different level of study, namely at the nucleotide level. As the study in this thesis is at both these levels, it is important to consider the findings of studies on sequence analysis of histone genes. Publication of a wide range of histone gene sequences has allowed analysis of the fine structure of histone genes to be undertaken. These analyses have the potential to help expose regulatory sequences of a tissue-specific, gene-type specific, species-specific, or temporal kind. Furthermore they allow evolutionary studies to be performed (eg Busslinger *et al* 1982). Sequence comparisons were initially concerned with comparisons between the various sea urchin clones (reviewed Hentschel & Birnstiel 1981), but have been extended to many species, including *Xenopus* (Turner & Woodland 1982, Perry *et al* 1985). Almost all histone genes have no introns, and so are composed of a domain coding for the structural gene flanked by upstream 5' sequences, and downstream 3' sequences. The fine structure of these regions is considered in turn.

1.D.1 CODING DOMAIN

Many of the published histone gene sequences have undergone codon usage analysis. The results of these analyses indicate that within every species investigated there appears non-random codon usage (see Maxson *et al* 1983c). Analysis of

two species of sea urchin indicates that certain amino acids in each species display a codon bias, while other amino acids show equal use of the alternative codons. Interestingly, those amino acids with preferential codons in one species appear to display a similar bias in the other species. Furthermore, those lacking a bias in one species were also found to display no bias in the other species. Analyses of *Xenopus*, newt, and mammalian histone codon usage indicates a specific trend of under-use of U and A in the third base positions (Turner & Woodland 1982, Tichter & Graur 1989). A recent study (Hurt *et al* 1989) indicated that sequences within the coding region of a murine H3 gene play a role in gene regulation, as they are required for high level expression. The preference for a particular codon, could be explained by the requirement for the maintenance of such sequences.

1.D.2 REGIONS 5' TO CODING SEQUENCES

In this section the various characteristics of the 5' flanking region will be considered, starting with those sequences closest to the initiator codon.

INITIATION CODON BLOCK

The initiator codon forms part of the first homology block of sequence 5-'CAPyNATG-'. This consensus was originally thought to be sea urchin-specific (Busslinger *et al* 1980), however studies on sea urchin late genes (Childs *et al* 1982)

show this consensus not to be present. Thus, this sequence may play a role in regulation of expression of the early genes. It is unlikely that this sequence should be preserved between very distantly related species of sea urchin, without conferring some biological function, although, the evolutionary behaviour of other histone gene sequences between distantly related sea urchin species has proved difficult to understand (Busslinger *et al* 1982).

Comparison of the sea urchin consensus sequence with the corresponding region of a *X. laevis* H1 gene revealed only minor divergence (Turner *et al* 1983). However, more recently published *Xenopus* histone gene sequences display greater divergence from the sea urchin consensus sequence (Perry *et al* 1985).

CAP SITE

The cap site refers to the 5' end of the mRNA which occurs at the DNA sequence 5'-PyCATTCPu-3'. This motif has been identified from sequencing genomic clones (Sures *et al* 1978), and is similar to the sequence 5'-PyPyAUCPu-3' obtained from analysis of histone mRNAs (Sures *et al* 1980). One could envisage that transcription was either initiated at this point, or that processing of a long transcript to this point yielded the correct mRNA 5' terminus. Work on several species has shown that the former is the case (Sures *et al* 1980, Sugarman *et al* 1983, Seiler-Tuyns & Birnstiel 1981). More information on the function of the cap site has

come from microinjection of mutant genes lacking this sequence into *Xenopus* oocytes. Results indicated that spurious transcription started some 60 bp downstream relative to the intact gene (Grosschedl & Birnstiel 1980b). This shows that the cap sequence must be involved in the positioning of the start of transcription, rather than solely regulating the amount of transcript produced. In the sea urchin early histone genes, the cap site occurs 77 bp upstream from the AUG initiator codon. The histone genes are unusual as they do not display the common 5'-CTTPyTG-3' which occurs closely downstream from the cap site in most eukaryotes (Barralle & Brownlee 1978).

TATA

The 'TATA' or Goldberg-Hogness box, with the sequence 5'-GTATAAATAG-3' is probably the most commonly quoted 5' homology block, and is ubiquitous to eukaryotes. It occurs some 20-30 bps upstream from the cap site. This sequence has been demonstrated to be part of an RNA polymerase II promotor (Breathnach & Chambon 1981), and is consequently found preceding most eukaryotic RNA polymerase II genes (Busslinger *et al* 1980).

Deletion studies of regions containing the 'TATA' box using *in vitro* systems have shown that this sequence is the major determinant in promoting specific initiation of transcription (Corden *et al* 1980, Hu & Manley 1981, Tsai *et al* 1981, Wasylyk & Chamon 1981). Related *in vivo* studies

using the *Xenopus* oocyte microinjection system, showed that unlike the related prokaryotic Pribnow box, the removal of this sequence did not abolish all transcription, but caused a reduction to 22 % of the level produced by the unaltered sea urchin histone gene (Grosschedl & Birnstiel 1980b). As three discrete classes of mRNA were observed in its absence, the authors argue that this sequence 'selects' the correct 5' mRNA terminus. A point mutant change from 5'-TATA-3' to 5'-TAGA-3' showed a reduction of 80 % in transcription (Grosschedl & Birnstiel 1982). The role of the TATA box in transcription activation has more recently been established, following the isolation of the TFIID protein. This protein binds to the TATA box and initiates a cascade of assembly of several general transcription factors and RNA polymerase II, to form the pre-initation transcription complex (see Lillie & Green 1989).

UPSTREAM REGULATORY SEQUENCES (URSs)

Regulatory sequences in the 'far upstream' region also known as, 'modulators', or enhancers, occur 5' to the 'TATA' box, and are known to occur upstream of histone genes. *In vivo* studies using deletion mutants have shown these additional sequences to be necessary for maximal transcription (Hentschel *et al* 1980a). Mapping of regions from *P. milliaris* histone genes has shown that sequence elements between -139 and -111 bp upstream of the transcription start, enhance transcription. Interestingly this effect on

transcription can be mimicked to a degree by the presence of free DNA ends (Grosschedl & Birnstiel 1982). The authors suggest this may then act as an entry site for RNA polymerase II.

In this upstream region however, there are other regions of homology. Firstly, the sequence 5'-GATTC-3' has been shown to occur about 10 bp upstream from the 'TATA' box, and to be a histone-specific homology block (Sures *et al* 1978, 1980, Hentschel *et al* 1980). Secondly, the pentamer 5'-CCAAT-3', which is related to the homology block 5'-GGPyCAATCT-3' (Benoist *et al* 1980, Efstradiadis *et al* 1980), is found in the promoter region of many eukaryote genes, including histone genes. Several different factors have been shown to bind to this element (see Dorn *et al* 1987). One such recently described factor, H1TF2, activates transcription of a human histone H1 gene (Gallinari *et al* 1989). Several transcription factor binding domains are being uncovered in the upstream region. Tung *et al* (1989) describe a sea urchin H4 gene-specific transcription stimulating factor, which binds in the region -102 to -436. One factor of wide interest, is OTF1 (Fletcher *et al* 1987). This factor binds to the octamer motif ATTTGCAT, and stimulates the transcription of the human H2B gene. Interestingly, the octamer element also is found in promoters and enhancers of immunoglobulin and snRNA genes.

The importance of the conserved elements found 150 bps or more upstream, has not been fully determined, and appears to have been initially overlooked. Deletion studies of these regions, can cause a 15-fold decrease in transcription, (Grosschedl & Birnstiel 1980b), which contrasts with the 5-fold decreases in transcription observed on deletion of other closer sequences, including the TATA box, and cap regions, under identical conditions. In this respect, URSs appear to differ from the TATA and cap regions which not only determine amount, but also effect correct initiation. It appears deletion in the far upstream regions cannot cause spurious initiation starts, possibly as it does not affect RNA polymerase II recognition sequences.

Work in yeast has identified a region upstream from the 5' region of an H2A histone gene (ie from the intergenic region between H2A and H2B) which contains two cell-cycle conferring elements (Osley 1986). One contains three repeats of a 16 bp element, involved in the periodic activation of transcription. The other is a 19 bp element displaying dyad symmetry, which appears associated with a cell cycle dependant repressor. The URS homology blocks, originally characterised in sea urchin, have prompted further study in other species. In general, the higher eukaryotes appear to display much more highly diverged prelude sequences. In the chicken H2B genes, regions of homology to the 'CCAAT', 'TATA', and 'cap' are apparent, but are found at different distances apart from those in all the

sea urchin species. There appears little other similarity between 5' flanking regions of the chicken and sea urchin, except for the 13 bp homology block, which is common to H2B genes. (Harvey *et al* 1982). *Xenopus* shows a similar degree of divergence, with only the 'CCAAT' and 'TATA' regions sharing sea urchin sequence homology. When considering the various homology blocks that have been identified and characterised, it is important to note that as well as distinct regions of conservation, there are other weak homology sequences which can be noted by either weak sequence homology, or by the fact that they display an overall high 'GC' content, which distinguishes them from the 'AT' rich spacer DNA (Shaffner *et al* 1978). This indicates that these are regions which may confer biological function.

1.D.3 REGIONS 3' TO CODING SEQUENCES

RNA molecules must obtain a 3' end by one of two methods, namely transcription termination, or processing of a longer transcript. Prokaryotes usually adopt the former strategy, whereas, eukaryotes adopt the latter. RNA polymerase II continues to synthesise RNA a considerable distance beyond the 3' ends of the mature mRNA, intramolecular cleavage followed by sequential addition of adenosines is then thought to follow (Darnell 1982). The components involved in this processing are being identified (reviewed Birnstiel *et al* 1985, Schumperli 1988).

POLYADENYLATION SIGNAL

Histone genes are unlike other RNA polymerase II genes in that the majority of histone mRNAs do not display a polyadenylation tail. The notable exceptions to this are of two types. Firstly the yeast histone genes produce fully polyadenylated mRNAs, which are like the typical RNA polymerase II genes in that polyadenylation occurs partly due to the presence of the sequence 5'-AAUAAA-3'. This sequence has been shown to be important in processing and polyadenylation of the 3' terminus of mRNAs (Fitzgerald & Shenk 1981). The motif is found at a relatively fixed distance from the 3' end of the mature mRNAs, and is highly conserved (Montall *et al* 1983)

The second group are those histone genes which do not display this signal sequence, yet appear to produce polyadenylated mRNAs, although the length of these tails may be shorter than for genes which do display this signal. Examples of this type are the histone mRNAs in the ovary of *Xenopus*, and other selected genes, eg. the chicken H5 gene and chicken H3.3.

HYPHENATED DYAD SYMMETRY BLOCK

A 23 bp homology block with the sequence 5'-AACGGCc/tCTTTTCAGg/aGCCACCA-3' has been found to be present at the 3' end of histone genes. It has been observed in every species studied except yeast (Birchmeier *et al*

1982). Within this sequence is a 16 bp hyphenated inverted repeat, which can react intramolecularly to form a hairpin loop structure. The sequence, not just the secondary structure, has been conserved with great fidelity over a long evolutionary time span (reviewed Mentschel & Birnstiel 1981), which led to the proposal that it represents a regulatory signal in either transcription termination (Busslinger *et al* 1979) or nucleocytoplasmic transport (see Mentschel & Birnstiel 1981). Mutants lacking only 12 bp within the inverted repeat were microinjected into *Xenopus* oocytes, under conditions that allowed production and identification of correctly sized sea urchin histone H2A mRNA. This mutant gene produced no correctly sized mRNAs, but S1 protection experiments displayed correct initiation was occurring. Furthermore this mutant produced the expected level of transcription, and the stability of the transcripts was unaffected. This showed a variety of 3' termini existed, many of which extended past the deletion site into the downstream spacer. Defective 3' processing of a sea urchin H3 gene, when injected into *Xenopus* oocytes, was found to be rectified by injection of a chromosomal wash from sea urchin (Stunnenberg & Birnstiel 1982). The responsible factor was found to be a 60 bp RNA molecule, termed U7 (Galli *et al* 1981). The U7 sequence has regions complementary to the part of dyad element, and also a second homology block (see below). Although not fully understood, it seems likely that pairing between the U7 and 3' element does play a part in

processing (Strub *et al* 1984)

Neither this 23 bp sequence, nor any like it is found in the 3' region of the chicken H5 histone gene (Kreig *et al* 1983).

3' TERMINUS

The location of the 3' terminus of histone mRNAs has been determined by S1 mapping experiments, which have shown it to occur near the 5'-ACCA-3' motif at the 3' end of the 23 bp homology block. (Hentschel *et al* 1980a, Sailer-Turns & Birnstiel 1981). This is in agreement with sequences of *Xenopus* histone cDNA clones (Zernik *et al* 1980, Turner & Woodland 1982). Thus in this respect eukaryotes are similar to prokaryotes; prokaryotic transcripts terminate in a hairpin structure (reviewed Pribnow 1979).

SECOND BLOCK

A second homology block, only a few (7-10) nucleotides downstream from the dyad element has been discovered which again is found to be conserved in all histone genes studied, although not with the fidelity of the 23 bp block (Turner & Woodland 1982, Turner *et al* 1983). The vertebrate consensus sequence is 5'-AAAAGACCTGA-3'. This second block is thought to base pair with a region of the U7 snRNA, and together with a second component, regulate the histone 3' processing (reviewed Schumperli 1988).

1.D.4 CONCLUSION

In this introduction I have summarised information on histone gene structures and arrangements over a wide range of species. The work described in this thesis contributes to this body of information by considering the histone gene arrangements of *X. borealis*. From section A a general trend becomes apparent. In the case of embryonic histone genes only, the invertebrates display a highly uniform, homogeneous histone gene cluster structure, while the warm-blooded vertebrates have highly heterogeneous gene arrangements. The late or adult form histone genes appear heterogeneous in all species. The arrangements in *Xenopus* appear to represent a form intermediate between those exhibited by the 'primitive' and more 'advanced' eukaryotes. In fact there are elements of both extremes in *Xenopus*.

One feature associated with histone genes in almost all species studied, is that of clustering. The reasons why histone genes are clustered at all remains unclear. Presumably, the factors determining cluster structure are a combination of evolutionary processes and functional selection. Old and Woodland (1984) make the point that the quintet arrangement, containing one copy of each histone type, is only found in organisms displaying a high histone gene copy number. Over a period of evolution, such an arrangement would be favoured for amplification if there was a demand for more copies, as its duplication would result in equal numbers of each type. Events

such as the genome duplication (eg as in *Xenopus*) would be an example where there would be an increased demand for histones in equal numbers, and thus it is not surprising that the majority of histone genes in *X. borealis* are in this quintet form. The high level of homogeneity between the reiterated quintets within each species that display this feature also requires explanation, given the inevitability of mutation. Recombinational events, and most notably unequal cross-overs, are thought to play a significant role. This specific point will be returned to in detail in chapter 6, but is worth noting at this point, as such processes must have affected the overall evolution of histone genes. It has been suggested (Harvey & Wells 1984) that the change from a highly ordered to a less ordered histone gene arrangement, during evolution, is a consequence of a reducednumber of recombination events, resulting from a lower gene copy number. Changes in population size and genetic drift could also have played a part. It is also possible that the close proximity of the genes is functionally advantageous. It is well known that the histone genes are mainly expressed during the S phase of the cell cycle. Possibly there is a lower metabolic burden on making a single region of chromatin transcriptionally active, than five distantly separated regions. Indeed, it is interesting that the 'replacement' histone genes that have been studied, which are not expressed in a cell-cycle dependant fashion, are not arranged in quintet clusters. The region around the chicken H5 gene (some 41kb) has been shown not to react with any histone gene sequences (Kraig *et al* 1983).

Introduction

How the observed variety in cluster structures described in this section has arisen, is still an open question. Possibly the different structural arrangements have been selected for by functionally significant features. As many of the different species display different developmental strategies, their requirement for histones in early embryogenesis varies. This could be a selective force to account for the differences in gene numbers (Birnstiel *et al* 1979, Woodland 1982). This can not however, account for all the differences in cluster structures described. Clearly the differences within *Xenopus* do not result from different developmental requirements. *Xenopus* provides a good example where there exist a wide range of cluster structures, and so a potential system for studying changes in cluster structure. The initial finding that *X. borealis* has a single predominant cluster type, while *X. laevis* appears not to, is most striking in this regard. Thus by characterising in detail the histone gene clusters in *X. borealis*, it may be possible to throw light on the significance of, and the mechanisms involved in, the evolution of histone gene arrangements over the broad spectrum of organisms described in this section.

The *Xenopus laevis/borealis* comparison has the potential advantage over other species considered in this section, that these two species are closely related. Evolutionary pathways may be easier to decipher over a shorter time span.

CHAPTER 2. MATERIALS AND METHODS

2.A. MATERIALS

Restriction enzymes and DNA modifying enzymes were obtained from the following companies and were used according to the manufacturers instructions: Boehringer Mannheim (London) Ltd, Bethesda Research Ltd, N.B.L Enzymes Ltd.
All radioisotope containing reagents were obtained from Amersham International plc, at the following specific activity: α-32p-dNTPs 2000-3000 Ci/mM at 10 mCi/ml in water. L-(4,5-3H) Lysine HCl 79Ci/mmol at 1mCi/ml. [35S] dATPαS 1000Ci/mmol at 10mCi/ml. Bacto-tryptone and yeast extract, for bacteriological media were obtained from Difco Laboratories (Michigan, USA) and Oxoid Ltd (UK).
X-ray film was obtained from Fuji Ltd. Photographic developer and fixer were obtained from Fuji, Eastman Kodak, or Ilford, and were used in accordance with the manufacturers instructions. NZ-amine was a gift from Dr. R.W. Old, and was originally obtained from Humko Sheffield Chemical Division of Kraft, Inc. 1099 Wall St. West Lynnhurst NJ 07071.
All other chemicals and reagents were obtained from BDH Chemicals Ltd (UK), Fisons Scientific Apparatus (UK), Sigma (UK), and Eastman Kodak (USA).

Bacterial and Phage genotypes

E. coli Strains JM103 : Delta(lac pro), thi, strA, supE, endA, sbcB, hsdR-, F'traD36, proAB, lacIq, zm15 (Messing *et al* 1981)

LE392 : F-, hsdR514(rk-mk-), supE44, supf58, lacY1 or delta(lacIZY)G, galK2, galT22, metB1, TrpR55, (Murray *et al* 1977)

K803 : hsdR-,hsdM-, gal-, met-supE, mcrA-, mcrB- (Wood 1966)

Phage vectors

Lambda 47.1 (Loenen & Brammer 1980); Lambda EMBL3 (Frischauf *et al* 1983) M13mp11 (Messing 1983)

2.B. METHODS

Percentage values indicate grams of substance per 100 mls water.

2.B.1 PREPARATION OF ERYTHROCYTE *XENOPUS* DNA

This work was performed in association with Prof. H.R. Woodland. The adult frog was anaesthesised in MS222 (0.2 % in water) for 20 mins, and then had the blood removed by heart puncture. N.M.T. + heparin <N.M.T.= 0.1M NaCl, 10mM Tris.Cl pH8.0, 1mM $MgCl_2$> was mixed with the collected blood, and used to facilitate extraction. The cells collected were washed twice in N.M.T + heparin followed each time by brief centrifugation at 3000 rpm for 5 mins at 0^{o}C. in a MSE Mistral 6L centrifuge fitted with a 8 x 50 rotor. A final repeat wash was performed as above but in the absence of

heparin. The cells were resuspended in N.M.T. + 1% NP40, and were left on ice for 10 mins. Centrifugation at 3500 rpm for 5 mins at 0°C. was used to pellet the nuclei, which were resuspended in 15 mls T.N.E.<10mM NaCl, 10mM Tris.Cl pH8.0, 10mM EDTA> and then had RNAase A to a final concentration of 40µg/ml added before incubation at 37°C. for 60 mins. Proteinase K to a final concentration of 40 Units/ml and SDS to 0.5% were added and the mixture was incubated, shaking at 37°C. overnight. Two phenol extractions were performed, with very gentle mixing, followed by a gentle chloroform extraction. The aqueous phase was transferred to a conical flask, and after addition of 1 /10 volume Sodium Acetate pH7.2 and two volumes of absolute ethanol, the DNA was spooled on a glass rod. After excess liquid had been removed, the DNA was redissolved in 15 ml T.E.<10mM Tris.Cl pH8.0, 1mM EDTA> at 4°C. for at least 48 hours.

2.B.2 DIGESTION AND GEL ISOLATION

Restriction digestions were carried out according to the suppliers instructions. The reaction was stopped by addition of 'Stopper dye' <30% Glycerol, 10mM EDTA pH8.0, Orange G to colour>. 20 µl aliquots were then loaded onto a 0.7% (or 0.4%) agarose gel (100ml, 15 x 20 cm). Agarose slab gels were run in TAE <TAE = 40mM Tris Acetate pH8.3, 1mM EDTA> (Murray & Murray 1974). To visualise the DNA, ethidium bromide was added to a final concentration of 1 µg/ml to both gel and buffer. Sizes of fragments detected in agarose

gels were calculated based on mobilities of known marker sizes. Size markers Lambda cI857S7 Hind III (Sanger *et al* 1982), and Lambda 47.1 EcoR1 (see Maniatis *et al* 1982) were used for larger fragments (0.6-24 Kb). Hinf1 digest of PAT153 (Sutcliffe 1978), were employed for the range 150 bp to 1.6 kb. The isolation of a particular fragment, after electrophoretic separation was performed as follows. A razor cut was made in the agarose gel, just on the positive terminal side of the band. Whatman 3MM paper, with a backing of dialysis membrane was placed into this slot, such that on further electrophoresis, the DNA passed into, but not through, the 3MM. After this, the 3MM paper was removed, and the DNA extracted by centrifugation. This was achieved using a small (600 µl) eppendorf tube with a needle hole in the base, placed inside a larger (1.5 ml) eppendorf tube. The Whatmann paper was then rinsed for 5 mins. with :0.2M NaCl, 50mM Tris.Cl pH 7.6, 1mM EDTA. The DNA was then precipitated by the addition of both 1/10 th volume 3M Na Acetate, and two volumes absolute ethanol, followed by one hour at -70^{o}C. The pellet was collected by centrifugation in an eppendorf microfuge for 10 mins. The pellet was drained, and dried at room temperature for 5 mins. under vacuum, before being redissolved in T.E. to the appropriate volume or concentration.

2.B.3 PARTIAL DIGESTS

10 µg of *Xenopus* DNA was added with restriction enzyme buffer in a final volume of 150 µl, and the contents were mixed well by inverting the tube several times. A 30 µl aliquot was dispensed into an appendorf tube (tube 1), and 15 µl aliquots likewise dispensed into tubes 2-8. The remainder was dispensed into tube 9, and all the tubes were chilled on ice. 8 units of restriction enzyme was added to tube 1, and was mixed well, to produce a concentration of 4 units/µg DNA. 15 µl of reaction mixture was transferred to tube 2, producing a concentration of 2 units /µg DNA in tube 2. The mixture was again mixed well, and the two-fold serial dilution continued through to tube 8. Nothing was added to tube 9. Tubes 1-8 were placed in a 37^{o}C. waterbath for an hour, and the reaction was stopped by the addition of 'Stopper dye'. The DNA was electrophoresed in a 0.4% agarose gel at 40-50 mA for 12-18 hours, in the presence of suitable size markers, and was stained with ethidium bromide.

2.B.4 LARGE SCALE PREPARATION OF PARTIALLY DIGESTED DNA

Bulk digestion of *X. borealis* DNA (250-500 µg) was carried out under conditions optimal for large insertions into each particular lamboid vector. An aliquot of this bulk digest underwent electrophoresis to check the fragment size range. For the Lambda 47.1 library, the different size fractions were separated on a 38 ml 10-40% sucrose density gradient, and were centrifuged in a Beckman SW27 rotor. The sucrose

buffer contained : 1M NaCl, 20mM Tris.Cl pH8.0, 5mM EDTA. Centrifugation was performed for 24 hours at 20°C. 0.5 ml fractions were collected, from which 10 µl aliquots were mixed with Stopper Dye and electrophoresed on a 0.4 % agarose gel. The 15-20 Kb size range gradient fractions were pooled, dialysed and precipitated.

2.B.5 PREPARATION OF LAMBDA 47.1 LIBRARY

Both ligation tests to determine optimum vector insert ratios, and the packaging reactions were performed by Dr. P. C. Turner. The titre of bacteriophage particles from the packaging reactions was performed as described below (see sections 2.B.9 and 2.B.10)

2.B.6 PREPARATION OF HYBRIDISATION PROBES

Three probes were prepared:

(a) SH115 -Major cluster Subsequent to a SstI/HindIII double restriction digestion of the clone XbMWJ02, a 0.75 Kb DNA fragment was gel isolated, and ligated into M13mp11.

(b) PH270 -Minor cluster Subsequent to a HindIII/PstI digest of the subcloned HindIII fragment displayed in Figure R17, a 270 bp fragment was gel isolated, and ligated into M13mp11.

(c) H1 gene -Minor cluster This probe was prepared in the same fashion as PH270, except that a HindIII/BamH1 double digest was used, and a 550 bp fragment was collected.

2.B.7 NICK TRANSLATIONS

The following reaction mixture was routinely used: DNA (usually 100-500 ng) in TE 5µl, 5xNick Translation Buffer (NTB) 4µl, TTP (10mM) 1µl, dATP (10mM) 1µl, α-32P-dGTP 0.5µl, α-32P-dCTP 0.5µl, DNAase (approx. $1x10^{-6}$U.) 2µl, DNA polymerase 1 (at 5U/µl)[Kornberg polymerase] 1µl, Water to a total of 20µl <NTB= 0.5M Tris.Cl pH 7.2, 0.1M $MgSO_4$, 1mM DTT, 500µg/ml BSA (fraction V)>. The mixture was then incubated at 15°C. for 3 hours. The reaction was stopped, and protein removed by a single extraction with an equal volume of equilibrated phenol. The aqueous upper layer was passed down a Sephadex G50 column (total vol.of 3.6 ml). Three drop fractions were collected, and those carrying the 'front peak' of incorporated radionucleotide were pooled. Incorporation of between 1 $x10^6$ and 1 $x10^7$ per µg DNA was usually obtained.

2.B.8 BACTERIOPHAGE PLATING

A bacterial colony was picked off an agar plate <1.5% Agar in L Broth>, with a sterile toothpick, and grown overnight in 10 ml of sterile L broth <1% Bacto tryptone, 0.5% Yeast Extract, 1% NaCl>. The bacteria were then harvested by centrifugation (1000g for 10 minutes) and were resuspended in 2 ml of 10mM $MgSO_4$. For each plate, 100 µl of bacterial suspension were mixed with 100 µl phage suspension and were allowed to preabsorb for 20 minutes at 37°C. before being added to 3.5 mls top agar at 45°C., mixed, and poured onto

prepared LB agar plates. These plates had been previously poured and dried in a plate dryer for 20 mins. Plaques became visible after incubation overnight at 37°C.

2.B.9 TITRATING THE LIBRARY

The libraries were titrated initially to determine optimum density of plaques on a plate. This was achieved by serially diluting the stock phage suspension in SM buffer <0.1M NaCl, 8.1mM $MgSO_4.7H_2O$, 50mM Tris.Cl pH7.5, 0.01% Gelatin.>, followed by plating as described above. The dilution which then yielded separate distinct, but closely spaced plaques was used in the plaque lifts. Sufficient clones were plated, at optimum density, to expect about 10 major cluster containing clones. This was based on the copy number determined by Turner and Woodland (1983).

2.B.10 PLAQUE LIFTS (Benton and Davis 1977)

The plates were cooled to 4°C. for 15 mins. before the plaque lift. Autoclaved squares of nitrocellulose (Schleicher & Schuell ltd, FRG) were laid on the plates for 2 mins and had three location marks made by vertical stabbing with a needle. The filters were then placed for 2 mins. in each of the three following solutions respectively: 1) 0.5M NaOH, 1.5M NaCl. 2) 0.5M Tris.Cl pH7.0, 3M NaCl 3) 2 x SSC <20xSSC = 3M NaCl, 0.3M Na Citrate, pH7.0> The filters were then air dried, and baked at 80°C. for 2 hours under vacuum.

2.B.11 HYBRIDISATION CONDITIONS

Filters from both plaque lifts and Southern blots were treated in a similar fashion. The baked filters were sealed in plastic bags, with 30 mls of prehybridisation buffer <100 mls buffer = 75 mls H_2O , 15 mls 20 x SSC, 20 mg PVP, 20 mg BSA, 20 mg Ficol> plus 1/10 volume poly adenylic acid (5')[1mg/ml]. The filter was then preincubated at 65°C. for at leasst 2 hours, after which time the filter was removed to a fresh bag with fresh buffer. The nick-translated probe was boiled for 10 mins., quenched rapidly in iced water, and added to the bag. The bag was sealed, and hybridisation allowed to occur overnight at 65°C. The filters were then washed several times in 2xSSC at 50°C., until the negative control (autoclaved nitrocellulose) did not register above laboratory background on the Mini-monitor (hand held bench Geiger counter). In the case of plaque lifts at this stage, positive controls (high density plating of the parent clone XbHW302) usually indicated a signal several fold above laboratory background. The filters were then air dried for several hours, and exposed to X-ray film, with the aid of an intensifying screen at -70°C. Typically plaque lifts were exposed for 18-36 hours, Southern blots for mapping, carrying recombinant DNA 2-12 hours, and Southern blots carrying genomic DNA for 12-24 hours.

2.B.12 PICKING PLAQUES

By making three location points with a needle, while the nitrocellulose was on the plate for the plaque lift (see section 2.B.10), and by dotting radioactive ink on these spots prior to autoradiography, it was possible to align the plaque and the 'positive' signal it produced. This done, a Pasteur pipette was used to remove an agar plug carrying the plaque. In the case of the first screen of the library, when it was difficult to distinguish the exact plaque, a larger plug containing several plaques was removed using a blue (1ml) Gilson tip, which had had the tip cut off with a razor blade. In each case the plug was added to 1 ml SM buffer for storage, and subsequent replating. When replating was performed, the plaque was left at least an hour to form a phage suspension in the SM buffer, before aliquots were removed and plated to find the optimum density for the next screen.

2.B.13 PHAGE STOCK PREPARATION

Once a single plaque had been picked, and shown on rescreen to produce only positive plaques, the original plaque was considered pure. Pure plaques were used to produce a high titre stock as follows. The plaque was plated out at a high density, so as to produce a 'just confluent' plate. 3 ml SM buffer was then poured onto the plate (9 cm. diameter petri dish), left 30 mins., then removed to 2 large Eppendorf tubes, and centrifuged briefly to pellet unwanted bacteria.

This SM 'wash' was then stored over 25 µl chloroform, and used as the high titre stock. Plating of serial dilutions indicated that the titre was generally in the range of $1x10^6$-$1x10^9$ pfu/µl (pfu = plaque forming units).

2.B.14 LAMBDA DNA PREPARATIONS

Although a protocol using confluent agar plates was tried, it was abandoned in preference for a quick, small scale liquid culture method.

SMALL SCALE

A 10 ml aliquot of L broth was inoculated with the host strain (LE392 for lambda 47.1, K803 for Lambda EMBL3 recombinants), and allowed to grow, shaking overnight at 37°C. An absorbance measurement was then taken and the bacterial concentration calculated, based on 1 unit O.D. 600 nm = $8x10^8$ bacteria/ml. Aliquots containing 5 x 10^9 bacteria were then centrifuged at 1600 g for 10 mins. at room temperature. The pellet was drained and resuspended in 3 ml SM. To several of these, different numbers of phage were added, such that a range of ratios from 1 $x10^2$ pfu to 1 $x10^{10}$ pfu per 1 $x10^9$ bacteria was set up for each clone to be grown. The suspensions were placed in a 37°C. water bath for 20 mins. to allow preabsorbtion, and were then transferred to prewarmed sterile 100 ml aliquots of NZCYM <1% NZ amine, 85mM NaCl, 0.1% Casamino acids, 0.5% Bacto-yeast extract, 8.1mM $MgSO_4.7H_2O$, pH7.5> in 250 ml conical flasks, and shaken at 37°C. overnight. Concomitant

growth of bacteria and phage resulted in lysis after several hours. The flask which produced optimal lysis was then taken, and the contents centrifuged in two 'Oakridge' tubes at 10,000 rpm for 15 mins in an MSE MS18 fitted with a 8 x 50 rotor. The clear supernatants were then pooled, and had NaCl added to a final concentration of 1 M. To this 1g/ml PEG 6000 was added and dissolved. This was then left for at least an hour on ice to allow phage precipitation. Centifugation followed, at the conditions previously employed. The PEG pellet was then dissolved in 3 ml SM, 3 ml of chloroform was added, and, after vigorous mixing, centrifuged at 1600 g for 10 mins to separate the phases. The aqueous supernatant, was stored, and used in the presence of RNAase (boiled, 20 mg/ml) for restriction digestions, mapping etc.

LARGE SCALE

The procedure above was followed, except that aliquots containing $1x10^{10}$ bacteria were used, and the preabsorbtion suspension was added to 1 litre of medium, in a 2 litre flask. Centrifugation of lysed cultures was performed in 250 ml buckets. The procedure continues after the PEG pellet stage as follows. The redissolved PEG pellet in 3 ml SM, which was not chloroform extracted, had 1.5 g CsCl added, and was placed onto a CsCl step gradient of densities 1.7, 1.5, 1.45 g/ml (Yamamoto *et al* 1970). This was centrifuged overnight at 30,000 rpm in a MSE MS65, with a 6 x 14 rotor. The phage band was collected with a fine needle and syringe.

The phage were mixed with CsCl/SM solution of density 1.5 g/ml and centrifuged as above for 24 hours. The phage band was similarly re-extracted. Caesium was removed by overnight dialysis against TE (pH 8.0). The following were added to the concentrations shown: EDTA 20 mM, Proteinase K 50 µg/ml, SDS 0.5%, prior to incubation at 65°C. for one hour. This mixture was then phenol extracted twice, followed by a single phenol/chloroform extraction. The DNA was then precipitated with ethanol as described above (see section 2.B.2).

2.B.15 RESTRICTION SITE MAPPING

The mapping of the restriction enzyme sites along the length of each clone was performed by single and double restriction digests, and Southern blots. Approximately 1 µg of DNA was digested in a 20 µl volume according to manufacturers instructions. These were then loaded onto a 0.7% agarose gel, and run for a total of 700mA hours. The gels were photographed and used in Southern blot analysis. Mobilities were calculated with reference to known size markers (see section 2.B.2).

2.B.16 SOUTHERN BLOTS (Southern 1975)

Gels were soaked for 15 mins. in 0.5M NaOH, 1.5M NaCl, were then rinsed in a small quantity of 3M NaCl, 0.5M Tris.Cl pH7.0, and then soaked for 15 mins in the latter solution. The gel was placed on a wick of Whatmann 3MM paper, above a

reservoir holding 300 ml of 20 x SSC (see section 2.B.10), and using cling film, a window was prepared, through which transfer would occur. A sheet of prewetted (2 x SSC) nitrocellulose was placed above, covered with Whatmann 3MM paper. Above this the contents of a box of tissues was covered in turn with a glass plate to spread the pressure of a weight placed on top. After overnight transfer, and marking the origin, the filter was baked at 80°C. under vacuum for 2 hours prior to prehybridisation. Agarose gels carrying high molecular weight DNA were placed for 15 mins. in 0.25M HCl prior to transfer.

2.B.17 LARGE SCALE PREPARATION OF M13 rfDNA AND PLASMID DNA

M13 rfDNA was prepared by the 'Cleared Lysate' procedure (Clewell & Helinski 1970) modified slightly. A white plaque was toothpicked into 1.5 ml 2xTY <1% Bacto-tryptone, 0.5% Yeast Extract, 1% NaCl> and grown shaking, for 6 hours at 37°C. Meanwhile an overnight culture of JM103 was added to 1 litre of prewarmed, sterile 2xTY and grown for 3 hours. The infected suspension was then added, and grown for a further 4 hours. The bacteria were harvested by centrifugation for 30 mins. at 2,500 r.p.m. in a Mistral 6L centrifuge fitted with a 4 x 1000 rotor. Each pellet was then resuspended in: 3mls 0.25M Sucrose, 2mM $MgCl_2$, and transferred to an 'Oakridge' tube. 1 ml of 40 mg/ml lysozyme in 40mM Tris.Cl pH8.0, together with 5 µl of boiled 20 mg/ml RNAase were added and the tube was left on ice for 5 mins. 1 ml of 0.25M

EDTA pH8.0 was added and the mixture left a further 5 mins on ice. The bacterial remains were removed by centrifugation at 24,000 r.p.m. for 30 mins. at 4°C. in a MSE HS65. 1/10 volume of Proteinase K (10 mg/ml) was added and the mixture incubated at 37°C. for 30 minutes. The solution was then extracted twice with equal volumes of neutral equilibrated phenol, and the DNA precipitated with 1/10 volume 3M Sodium Acetate pH7.2 followed by the addition of 2 volumes absolute ethanol, and left at -70°C. for 1 hour. In the case of the H4 gene plasmid, pcXLH4W2 (Turner & Woodland 1982), appropriate antibiotics were added at the time of host cell infection, and for amplification, 150 mg of solid chloramphenicol was added, prior to a further incubation, shaking at 37°C. for 14-18 hours (Clewell 1974). After purifying the DNA, the H4 gene was excised by a BamH1/EcoR1 double digest.

2.B.18 CAESIUM CHLORIDE EQUILIBRIUM CENTRIFUGATION

M13 rfDNA and plasmid DNA were purified by banding within a CsCl/ethidium bromide equilibrium gradient (Vinograd 1963, Radloff et al 1967). The DNA precipitate was obtained by centrifugation (3,000 r.p.m. in a Mistral 6L with a 8 x 50 rotor for 10 mins.), and drying the drained pellet under vacuum for 5 mins. at room temperature. This DNA was dissolved up in 20 ml T.E. <10mM Tris.Cl pH8.0, 1mM EDTA>, then added to 20g CsCl, producing a final volume of approximately 25 ml. This was transferred to a Beckman

'self-seal' centrifuge tube (38 ml), and 1.5 ml 10 mg/ml ethidium bromide was added. The remaining space was filled with light paraffin oil (BDH) then balanced with a 1g/ml CsCl/T.E. solution. Next the tube was heat sealed and centrifuged in a Beckman L8 centrifuge fitted with a VTI50 rotor for at least 16 hours at 45,000 r.p.m., and 18°C. The lower of the two bands visible under UV illumination was supercoiled M13 rfDNA or plasmid DNA. This band was removed using a needle and peristaltic pump. The ethidium bromide was then removed by repeated extraction with equal volumes of butanol previously equilibrated with water saturated with CsCl. Efficient separation of the partition layers, was aided by centrifugation for 5 mins. at low speed in a bench-top MSE centrifuge. Typically four or five extractions with equal volumes sufficed. The solution was then dialysed overnight in 1000 fold volume of T.E. (pH8.0), and precipitated with ethanol, as described above (section 2.B.2).

2.B.19 PREPARATION OF M13 VECTORS

DNA of M13 mp10 and mp11 was prepared in bulk as described above (section 2.B.17). 1 µg of each DNA was digested with the appropriate enzyme in a total reaction volume of 20 µl, under the recommended conditions. To this mixture the following were added: 10% SDS 0.5µl, 1M Tris.Cl pH 8.0 1µl, CIAP 2µl, H_2O 26.5µl. This mixture was incubated at 37°C. for 30 mins. The following were then added: 10% SDS 4µl, 3M

Na Acetate (pH7.2) 5 µl. The DNA was extracted with equilibrated phenol (pH7.0), chloroform, and ether extractions respectively, and was then ethanol precipitated and redissolved in 75 µl of T.E.

2.B.20 LIGATION

The 'target' DNA (to be inserted) was added to 20-50 ng of the M13 vector with 1 µl ATP (10mM) and 1 µl 10xC buffer <10xC = 0.5M Tris.Cl. pH7.5, 0.1M $MgCl_2$,10mM DTT>, with 1 unit of DNA ligase, in a final volume of 10µl. The reaction was allowed to proceed at 4°C. overnight. Ligation controls without target were routinely included.

2.B.21 TRANSFORMATION

JM103, from a single colony, was grown overnight shaking at 37°C. in 2xTY. 100 µl of this culture was added to a further 10 ml of 2xTY and grown shaking at 37°C. for a further 2 hours (O.D. 600 = 0.4-0.6 Units). The bacteria were harvested by centrifugation at 2000r.p.m. and 4°C. in a bench-top centrifuge. The pellet was resuspended in 3 ml TFB <10mM Mes pH6.2, 100mM KCl, 45mM $MnCl_2$, 10mM $CaCl_2$., 3mM Hexaminocobaltic Chloride> and left on ice for 15 mins. before centrifugation as above. The pellet was again resuspended in TFB, this time in 800 µl, to which 33 µl DMF was added. The bacteria were left 5 mins on ice after which 33 µl of DTT (2.25 M DTT, in 40mM K Acetate pH6.0). After a further 10 mins on ice, another 33 µl of DMF was added,

rendering the bacteria competent after a further 5 mins on ice. 200 µl of these bacteria were added to the ligation mixture, and left for 40 mins on ice prior to plating. Plating 3.5 mls of molten H top agar was aliquoted into prewarmed tubes at 45°C. and was allowed to equilibrate. The transformation tubes were flicked and placed at 45°C. for 3-5 mins. Meanwhile 25 µl of BCIG (24 mg/ml in DMF), 25 µl IPTG (24 mg/ml in H_2O), and 200 µl of exponentially growing JM103 were added to the agar. The heat shocked transformation mixture was then also added, mixed, and poured onto prepared plates.

2.B.22a SEQUENCING GELS

Gradient gels were prepared as follows: Raven gel plates were washed well with warm water and polished with ethanol. The notched front plate was siliconised by use of Dimethyldichlorosilane solution (2% in 1,1,1-trichloroethane). The gels were taped using X-ray film spacers. The gel components were prepared as follows. (a) 2.5 x TBE <TBE = 45mM Tris borate pH8.3, 1mM EDTA>, Acrylamide 2.7g, Bis-acrylamide 0.135g, Urea 18.9g, Sucrose 2.5g, Bromophenol Blue to colour - made up with 12.5 ml 10 x TBE and H_2O to 45 mls. (b) 0.5 x TBE, Acrylamide 2.7g, Bis-acrylamide 0.135g, Urea 18.9g. - made up with 2.5 ml 10 xTBE and H_2O to 45 mls. 12 ml of (a) and (b) each had 90 µl 6.6% AMPS and 14 µl TEMED added, and were separately mixed. A 25 ml pipette was first filled with 12 mls of (b) followed

by 7 mls of (a). Limited mixing was allowed by sucking up about 10 air bubbles through the mixture. This 'gradient' mixture was then released into the pretaped gel plates, and a 12 place sharkstooth 10/1000 inch spacer was inserted. The gels were left at least an hour, but routinely overnight, to fully polymerise before use. The gels were run at a constant current of 27 mA with an upper reservoir of 0.5 x TBE, and a lower reservoir of 1 x TBE. The gels were fixed for 15 mins. in 10% Acetic Acid, 10% Methanol, then dried at 60°C. onto the back plate for 2-4 hours, and lastly exposed to X-ray film overnight.

2.B.22b SEQUENCING REACTIONS

(a) Annealing: The following annealing reaction was used: ssDNA 2µl, 10 x Hin Buffer 1µl, M13 primer (2.5 ug/ml) 1µl, H_2O 6µl, <10xHin buffer = 0.1M Tris.Cl. pH7.4, 0.1M $MgCl_2$, 0.5M NaCl>. Annealing was carried out by immersion in a 90°C. waterbath for 3 mins. followed by cooling for 15 mins. at room temperature.

(b) Deoxy-Dideoxy mixtures: Ratios of N^0 and ddNTP solutions (10mM, in 5mM Tris.Cl. pH8.0, 0.1mM EDTA) were altered to give optimum sequence. A^0 mixture contained 0.066 mM dGTP 1µl, 0.5 mM dCTP 20µl, 0.5 mM dTTP 20µl, 0.5 mM dATP 1µl. The other N^0 solutions likewise had the reduced volume of the respective dNTP. The G^0 solution had 20µl of each of the other three dNTP solutions.

(c) Reaction mixtures: The Klenow polymerase (5 U/µl) was diluted 1 in 4 with α-^{32}PdGTP, and Klenow dilution buffer <50mM K_2HPO_4 pH7.0, 2.5mM DTT, 50% glycerol), to a final volume of x µl, where x equals the number of samples times 1.5µl. 2µl of M^0 plus ddNTP was added with 1.5µl reaction mixture to the annealed mixture. The reaction was allowed to proceed for 15 mins. at room temperature. The 'chase' reaction was started by the addition of 2µl of 0.5mM dGTP, and was also allowed to proceed for 15 mins. The reaction was stopped by the addition of 4µl 'Stop' dye <0.3% Xylene cyanol, 0.3% Bromophenol Blue, 10mM EDTA pH7.5, made up in deionised formamide>.

2.B.22b EXTENDED SEQUENCING

The following solutions were prepared: Extension mix (5x); 7.5µM dCTP, 7.5µM dGTP, 7.5µM TTP. Termination mixes A; 300 µM ddATP, 25 µM dATP, 250µM dCTP, 250µM dGTP, 250µM TTP, C; 100 µM ddCTP, 25 µM dCTP, 250µM dATP, 250µM dGTP, 250µM TTP G; 150 µM ddGTP, 25 µM dGTP, 250µM dATP, 250µM dCTP, 250µM TTP, T; 500µM dTTP, 25µM TTP, 250µM dATP, 250µM dCTP, 250µM dGTP, Reactions: Following standard annealing procedure (above), Tubes labelled A,C,T,G, were filled with 2.5µl of appropriate termination mix. To the annealed reactions, the following was added: 2µl Extension mix (1x 1µl α^{35}dATP, 1µl 0.1M DTT, 2 Units Klenow Enzyme, Water to 6µl. This mixture was mixed by brief centrifugation, and incubated for 5 mins at room temperature. 3.5µl of this

extension reaction was then added to each of the four tubes labelled A, C, G, and T. These were spun briefly in the microfuge, and incubated for 5 mins at 37°C. This reaction was stopped as in the standard procedure, and samples processed further in the same way. Gels were run under varying conditions, but such that the temperature never rose above 50°C.

2.B.23 MICROINJECTION OF HISTONE CLONES

This work was carried out in association with Prof. H.R. Woodland. A female *Xenopus* frog was anaesthesised in MS222 (0.2% in water) for 20 mins. After removal of part of the ovary, the donor frog was stitched up again and returned to the laboratory stocks. The nuclear microinjection procedure was exactly as described in Old *et al* 1982. In short, stage 6 oocytes were separated from the surrounding cells and were used for microinjection. Samples of about 30 oocytes were taken. DNA samples of about 20-50nl were microinjected at a concentration of 0.5 mg/ml. in distilled water aiming for the nucleus.

Histone labelling:

Total proteins were labelled by incubating the microinjected oocytes in 5 mCi/ml Tritiated Lysine (L-<4,5-^{3}H> Lysine monohydrochloride) in Barth X at 21°C. overnight, in a humid atmosphere to reduce evaporation. Barth X = equal volumes of BXA and BXB. <25xBXA =2.2M NaCl, 2.5mM KCl, 6mM $NaHCO_3$, 0.375mM Tris.Cl. pH7.8>. <25xBXB = 8.2mM $Ca(NO_3)_2$.

10.3 mM $CaCl_2$, 20.3mM $MgSO_4$>. Incubated oocytes were removed singly, were punctured and squeezed to facilitate the removal of the germinal vesicle. The remaining cytoplasm was stored in Kressmans buffer <10mM Tris.Cl pH7.4, 1.5mM $MgCl_2$, 10mM NaCl>. The nuclei were added to carrier protein, prepared as total oocyte protein, from five oocytes. Each sample containing carrier protein was homogenised in 0.2M sulphuric acid, 100mM PMSF (Phenyl methane sulphonyl fluoride), to which 5 µl of 10 mg/ml Calf Thymus Histone was added. This mixture was spun in an eppendorf microfuge for 5 mins. The supernatant was removed and added to 10 volumes of absolute ethanol, prior to storage at -20°C.

Preparation for SDS gel electrophoresis: The sample was centrifuged for 20 mins at 2500rpm in a MSE Mistral 6L centrifuge fitted with a 8x50 rotor, and the supernatant was removed. The pellet was washed twice in 90% ethanol, to remove the sulphuric acid, prior to desiccation under vacuum. The lyophalised pellet was dissolved in 50 µl 'Reduction Cocktail' of: 10 M Urea (Ultra pure), 20mM NH_3HCO_3, 5% beta-mercaptoethanol. 15µl of this was added to 15µl of 2x SDS sample buffer, to be loaded on the gel. <2 x SDS Sample Buffer= 20% Sucrose, 2% SDS, 5% beta-mercaptoethanol>.

2.B.24 SDS GEL ELECTOPHORESIS

18% Acrylamide SDS electrophoresis gels were prepared as follows: 100 ml of separating gel mix was produced by the addition of the following ingredients: 56.2 ml H_2O, 12.5 ml Tris.Cl buffer <3M Tris.Cl pH8.8>, 1.0 ml 10% SDS, 20µl TEMED, 0.5ml AMPS. This mixture was degassed for 10 mins. by stirring while under vacuum, prior to pouring. The gel plates were 20 x 48 cm., prepared with 1 mm greased spacers. Distilled water was placed above the main separating gel prior to polymerisation. Polymerisation was allowed to proceed at room temperature for at least 90 mins. The water was then poured off. 20 mls of stacking gel mix was prepared by the addition of the following ingredients : 66 ml Stacking Acrylamide <10.0g acrylamide, 0.5 g bis-acrylamide, in 100 ml water>, 4.8ml Stacking gel buffer <0.5M Tris.Cl pH6.8, 0.2ml 10%SDS, 10µl TEMED, 0.2ml AMPS>. The stacking gel was again left at least 90 mins prior to running the gel. The gel was run at 40 mA until the samples had run into the stacking gel, then overnight at 10-14 mA in SDS gel running buffer.<0.4M Glycine, 50mM Tris.Cl pH 8.5> 1.5µl and 2.5µl of cytochrome C at 10 mg/ml were loaded to allow visualisation of the extent of electrophoresis. The gel was stained in: 0.1% Coomassie Brilliant Blue, 45% Methanol, 10% Acetic acid, in water, for one hour. The gel was destained in: 45% Methanol, 10% Acetic acid, in water, for at least one hour.

2.B.25 TRITON/ACID/UREA (TAU) GELS

Triton/Acid/Urea gels for histone separation were produced as follows: 80 mls of gel mixture was prepared by the addition of the following; Acrylamide (stock 60%/0.4% bis) 16.0 mls [12%], Urea 36g [7.5M], Acetic acid 4mls [0.9M], Triton X-100 (25% w/v stock) 1.2 mls [6mM], TEMED 0.4mls, H_2O to 79.5mls, AMPS (10% in H_2O) 0.48mls. The mixture was degassed for 10 mins., by stirring while under vacuum prior to addition of the AMPS. The mixture was then poured into prepared raven plates, with greased spacers, and left at least two hours to polymerise. The gel was pre-electrophoresed at 400 V overnight after the sample loading wells had been filled with Triton/Acid/Urea sample buffer <5% Acetic Acid, 7.5M Urea, 6mM Triton X-100>. A second, and third pre-electrophoresis steps were included, initially with the addition of 0.5 M Cysteamine (beta-mercaptoethylamine), 1mM DTT, and Methyl green to the TAU sample buffer. The aim of this was to scavenge for peroxides in the gel. The final 40 mins. pre-electrophoresis step was with the addition of protamine sulphate (0.5 mg/ml), to TAU sample buffer.

2.B.26 AUTOFLUOROGRAPHY

After destaining, gels were drained thoroughly, and fluorographed as follows: The gels were washed in 500 mls DMSO 1 for 30 mins, shaking gently, followed by an identical treatment in DMSO 2. The gels were then soaked for 3 hours

in 22% PPO (w/v) in DMSO. After draining, and a one hour wash in distilled water, the gels were dried onto Whatmann 3MM paper at 80°C. for an hour, in a vacuum gel drier, and exposed at -70°C. DMSO 1 is DMSO previously used as DMSO 2 five times. DMSO 1 maybe be used five times before being discarded. DMSO 2 is fresh DMSO. DMSO 2 maybe used five times before becoming DMSO 1.

2.B.27 CORRECTION FOR MULTIPLE BASE SUBSTITUTIONS

The following formular was used to correct for multiple base substitutions;

$$K = -3/4 \ln (1 - 4/3\Omega)$$

where K equals the corrected value, and Ω equals the uncorrected sequence divergence (Jefferies 1982, Kimura 1977).

CHAPTER 3. ANALYSIS OF THE *X. BOREALIS* MAJOR CLUSTER

3.A GENERAL INTRODUCTION TO RESULTS*

Prior to the work presented in this thesis, preliminary investigations on *X. borealis* had indicated that there is a major homogeneous histone gene cluster (Turner & Woodland 1983). Using genomic DNA, its gene order was determined and certain restriction sites in the genic region mapped. The number of H4 genes was measured as 80-90, about 70 % of which were present on the major cluster.

In this study the histone genes of *X. borealis* were investigated using both cloned and genomic DNA. The clones analysed can be divided into two classes based on cluster structure; these are the 'major' and 'minor' types, which will be considered separately in chapters 3 and 5 respectively. Analysis of the genomic organisation of the major type will be considered in chapter 4.

* The cloning, isolation, and initial characterisation of all the major cluster clones except XbHW9, XbHW11, and XbHW102 was performed in association with Dr. P.C. Turner. All the detailed mapping and most of the sequencing of clone XbHW302 was also performed in association with Dr. Turner. About one third of the sequencing of clone XbHW302 was performed exclusively by Dr. Turner, although all the analyses are my own work. The microinjection experiments were performed with the additional assistance of Prof. H.R. Woodland, and in some instances Ms J.E.M. Ballantine.

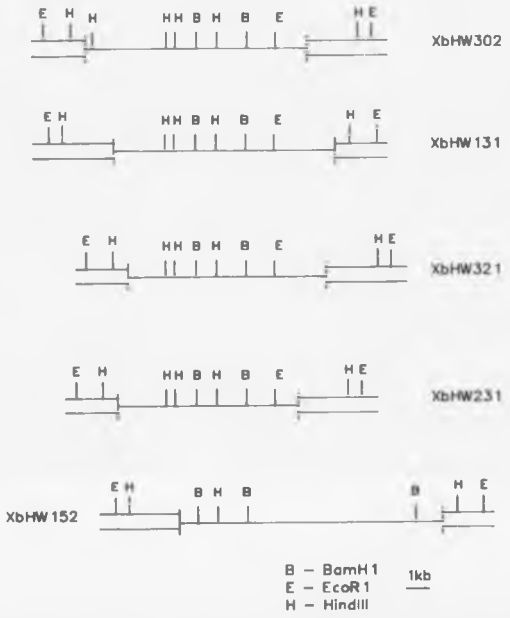


Figure R1 Rudimentary restriction site maps for the initially isolated major cluster clones.

XbHW152 contains additional unmapped EcoR1 and HindIII sites

3.B CLONING AND INITIAL CHARACTERISATION OF *X. BOREALIS* CLONES

In order to undertake detailed analysis of the histone genes of *X. borealis*, clones containing these genes were required. To this end a library in lambda 47.1 containing *X. borealis* genomic DNA partially digested with Sau3A was prepared. A mixed hybridisation probe of H4 and H1 *X. laevis* genes was used to identify histone containing recombinants. The screening yielded 36 initial positives, which were picked and rescreened. Twelve were screened to purity, and DNA was prepared for further analysis. DNA from each clone was digested with three restriction enzymes separately, fractionated by electrophoresis, blotted, and hybridised with histone probes for each major histone gene type (results not shown).

This yielded information which was used to prepare rudimentary restriction maps. From these maps it became apparent that the restriction sites of certain clones occurred with identical spacing, and these clones were grouped together accordingly. One such group of clones displayed a restriction site arrangement which matched the arrangement described as the *X. borealis* 'major' cluster (Turner & Woodland 1983). These restriction maps are displayed in Figure R1

Comparisons of the five initial clones (XbHW302, XbHW321, XbHW231, XbHW152, XbHW131) showed that the restriction site arrangements were very similar. Only clone

XbHW152 displayed a different profile. Southern blot analysis using probes from the non-histone gene and of clone XbHW152 on genomic DNA indicated the presence of repetitive elements (results not shown). Thus, this kind of analysis has not made it possible to determine whether this clone exists as such in the genome, or whether it is the product of multiple fragment cloning. It is possible that the BamH1 site at the right hand end of the mapped histone region in figure R1 marks the point where two fragments have ligated, as BamH1 sites encompass the Sau3a recognition/cleavage site. However, to reconstitute this BamH1 site the nucleotide base flanking the Sau3A site in the non-histone fragment would have to be a deoxycytosine, which occurs at a low frequency. Alternatively, this clone could represent the end of a repeated cluster. Subsequently minor differences in detail were found between the other four clones, showing that they do not represent the repeated cloning of a single cluster. The high level of homogeneity is also consistent with the studies of *X. borealis* genomic DNA by Turner and Woodland (1983).

3.C CHARACTERISATION OF A MAJOR CLUSTER CLONE.

Clone XbHW302 was found to extend furthest to the left of the H1 gene as indicated in Figure R1, and was chosen for detailed study. Figure R2 shows a detailed restriction site map of clone XbHW302. The restriction site positions in this clone matched the previously published restriction map of the major cluster, based on direct genomic mapping (Turner & Woodland 1983).

To determine the nucleotide sequence of this clone, M13 subclones were prepared, containing fragments derived from the detailed restriction site map. The complete nucleotide sequence of all five histone genes, and certain flanking regions, was determined by use of the Sanger dideoxy sequencing method.

3.D SEQUENCE: MAJOR CLUSTER CLONE XbHW302

The sequence data is displayed in Figure A1 in the appendix. The key to the relative positions of this sequence data is given in Figure R2. Figure A2 in the appendix displays the predicted translation products of each of the five coding regions.

GENIC REGIONS

H1 gene

The major cluster H1 gene sequence can be compared by alignment to other published H1 sequences for *X. laevis*. Part A of Table (II) displays certain features of the

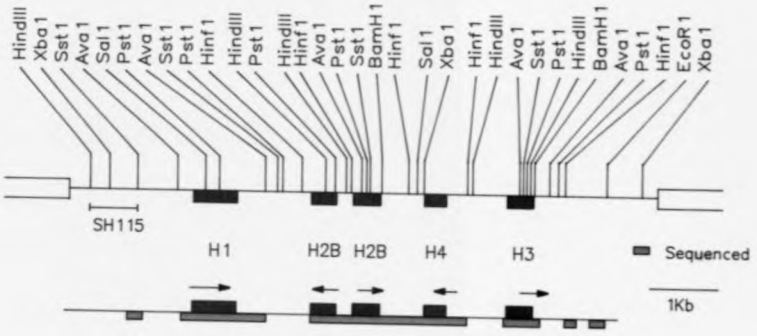


Figure R2 Restriction site map of clone XbHW302 displaying location and orientation of the histone genes, together with the extent of the sequence data displayed in figure A1.
The location of subclone SH115 is also displayed.

Results

<A> Subtype/source	H1A/Xlh3	H1B/Xlh1	H1C/XlHW2	XbHW302
Amino acid no.	211	221	218	217
Molecular weight	21376	22443	22348	22141
Acidic amino acids	10	8	8	9
Basic amino acids	59	66	67	66

Comparison with major cluster clone XbHW302 H1 gene.

<B> Amino acid level			
Matches (M)	183	179	172
Mis-matches	17	36	37
Unmatched	28	8	17
Length (L)	228	223	226
M/L %	80.3	80.3	76.1

<C> DNA level			
Matches (M)	540	525	502
Mis-matches	66	117	128
Unmatched	66	24	39
Length (L)	672	666	669
M/L %	80.4	78.8	75.0

Refs. H1A-Xlh3 Perry *et al* 1985
H1B-Xlh1 Perry *et al* 1985
H1C-XlHW2 Turner *et al* 1983

Table (II).

<A> Details of the three *X.laevis* histone H1 subtypes and the *X.borealis* major cluster histone H1 gene in clone XbHW302.

<B> Comparisons of the peptide sequences for the corresponding regions of the above genes.

<C> Comparison of the DNA sequence for the corresponding regions of the above genes. These comparisons are displayed in full in Figure A3 of the Appendix.

XbHW302 H1 sequence and certain published *X. laevis* H1 sequences. There is no clear indication from this data, which *X. laevis* subtype XbHW302.H1 is most similar to. Part B of this table displays a comparison using the alignment program of the Beckman Microgenie software package, to reveal the degree of similarity. The match/length percentage comparison analysis of subtype amino acid sequences does not clearly show whether the XbHW302 H1 gene is coding for an H1A or H1B subtype. It clearly is not similar to the H1C type. One of the differences between the H1A and H1B amino acid sequence is a 15 amino acid extension in the tail region of the H1B gene (Perry *et al* 1985). The H1A gene used in this comparison is unusual in lacking two highly conserved amino acid residues in the central conserved region, which are present in most other published H1 amino acid sequences (Lavy *et al* 1982). To determine whether this gene was of H1A or H1B type, a DNA sequence comparison was performed. This analysis indicates clone XbHW302 contains an H1A type gene. Two aspects of the analysis support this. Firstly the M/L % (match/length) figure is higher for the H1A. Secondly, H1B displays many more mismatched bases, while the H1A has more unmatched bases. All the analyses in Table (II) are displayed in full in Figure A3 in the Appendix.

Results

Upstream sequence elements

The sequences both 5' and 3' to the H1 gene were searched for homology blocks. Figure A1 (appendix) indicates the position of elements discussed below. The consensus 'Cap' site, PyCATTCPu, which was determined from studies on sea urchin histone genes (Busslinger 1980) was not found, but the hexamer 'GTGTTT' does occur in the region of the expected cap site. This sequence, which was noted by Perry *et al* (1985) is also found in the corresponding region of the H1 gene of *X. laevis* clone X1h3, and a related sequence is found in this position upstream of the H3 gene of X1h3 (Perry *et al* 1985). A 'TATAA' element occurs at -65 bp of XbHW302.H1, as may be expected. Interestingly 'CCAAT' elements are found both at -93 bp and -125 bp positions. Two 'CCAAT' boxes were found upstream of the H1 genes in the *X. laevis* clones X1HW8, and X1HW2 (Turner *et al* 1983), however in these clones, two 'TATA' boxes were also found. In contrast, clone XbHW302 displays the single 'TATA' box. The sequence data does not extend far enough upstream to determine whether the H1 histone gene class-specific elements described by Perry *et al* (1985) exist in clone XbHW302.

Downstream sequence elements

Both the 'dyad' and the second downstream homology blocks were identified, and their sequence and location falls within the variation displayed in the 3' sequence comparison made by Turner *et al* (1983).

Core nucleosome histone genes

All of the presumptive amino acid translation products indicate the expected amino acid sequence (figure A2).

At the DNA level, the sequence also displays typical features. As with the H1 gene, the sea urchin consensus 'cap' sequence was not found. For XbMW302.H3, there is insufficient data available to determine the presence or location of 'TATA' or 'CAAT' boxes. XbMW302.H2A, and XbMW302.H2B genes display the 'TATAA', motif at -73 bp and -79 bp respectively, while upstream of the XbMW302.H4 gene 'TATAT' occurs at -64 bp. These genes also display the 'CCAAT' motif at the following locations; -108 bp (H2A gene), -128 bp (H2B gene), and -155 bp(H4 gene). All of these motifs fall within the variation in location seen in previously published sequences, which are discussed elsewhere (see section 1.D). None of the genes display strong similarity to the consensus sequences of the histone class-specific elements proposed by Perry *et al* (1985). Sequences are present, however, which display a varying degree of similarity to the regions of clone X1h3 which are proposed to represent these class-specific

elements. Like the H1 gene, the 3' homology blocks were as expected, although there was insufficient sequence to identify the 3' elements of the H3 and H2B genes.

Previous analysis of H4 leader mRNA sequences in a variety of *X. borealis* tissues revealed a predominant class (Woodland *et al* 1984), which apart from length, varied only by a single base (see section 1.B.8c). The upstream XbHW302.H4 gene sequence is identical to one of these two sequences, and therefore shows that the predominant message class is indeed coded for by the major histone gene cluster. The presence of this mRNA class in all tissues and stages studied, reflects the ubiquitous expression profile for this cluster type. Since we had a further four examples of the major cluster, it was of considerable interest to determine just how variable their H4 genes were. The direct mRNA leader sequence was of course likely to show the population consensus. Clones XbHW231, XbHW321, and XbHW131 were therefore also sequenced in the H4 leader region. Like XbHW302, XbHW231 and XbHW321 displayed a deoxycytosine residue, while XbHW131 displayed a thymidine residue, at the exact location where the single base difference had been observed in the mRNA. Otherwise they were identical. This shows that both forms are coded for by almost identical copies of the major cluster. An idea of the high level of homogeneity between these four clusters is indicated by the fact that no other base changes were observed in 196 bps of non-coding sequence and 51 bps of H4 coding sequence.

Results

The sequenced genic regions of XbHW302 were compared, using the Microgenie software package, with the gene sequence of *X. laevis* clones Xlh1 and Xlh3 (Perry *et al* 1985). The percentage homology (match/length) of this analysis were as follows;

GENE	Xlh3	Xlh1
H2A	91.3	83.5
H2B	92.9	87.0
H3	92.0	84.7
H4	91.9	87.4

By way of example, the H4 gene comparisons are shown in full in the appendix figure A8(c) and (d). Clearly all the genes on clone XbHW302 are more closely related to those of Xlh3 than Xlh1. The implications of this are considered in chapter 6.

SPACER SEQUENCE

Non-repetitive sequences

The intergenic regions of clone XbHW302 were compared with both the fully sequenced *X. laevis* clones Xlh1 and Xlh3 (Perry *et al* 1985). The comparison parameters were set for a detection of greater than 80% match over any 10 bp. No significant homology (greater than 60 bp) was found in the comparison using clone Xlh1, but seven significant homology blocks were found when Xlh3 was analysed. The regions displaying homology are numbered 1 to 7 in figure R1. These

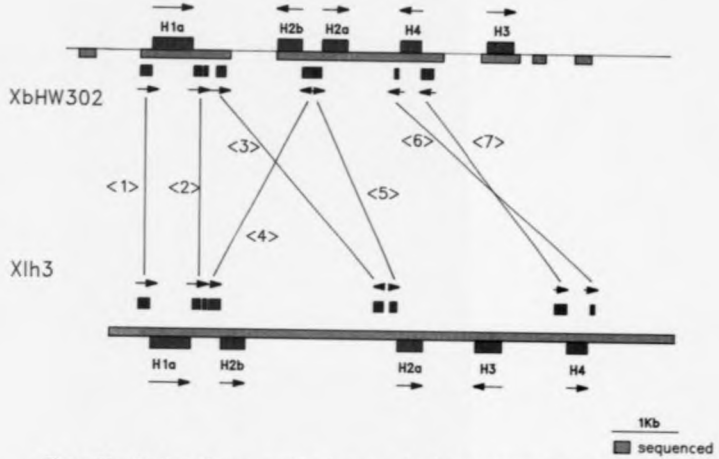


Figure R3 Regions of non-repetitive homology (red) of intergenic sequences between two histone gene clusters from Xenopus

analyses are displayed in figures A4.1 to A4.7 in the appendix. For each analysis, a matrix plot comparing the region of similarity from XbMWJ02 with the entire X1h3 sequence is presented, together with a sequence alignment of the homologous region. The matrix plot programme marks a dot, or short line, at the point where a preset degree of similarity is found between the two sequences represented by the ordinate and abscissa. Thus parallel lines represent the repetition of an element in one sequence, which is present once in the other (see A.4.6). Off-set lines indicate the translocation of an element (eg as in A.4.7). The matrix plot not only gives an idea of the overall homology pattern, but also shows that minor rearrangements have occurred within each region. This is indicated by the lack of a single continuous line. The DNA sequence below each matrix plot is of the longest region the computer could align, in each case.

Repetitive Sequences

As well as the above homologies, several regions of mono-, di-, and trinucleotide repeat elements were found to be common to clones X1h3 and XbMWJ02. These are shown in figure R4. Interestingly, the location pattern of these repetitive elements in these clones appears unrelated to the pattern of non-repetitive elements common to both (figure R3). This suggests that the evolution of the two sequence types may be subject to different forces. Such homocopolymers have been

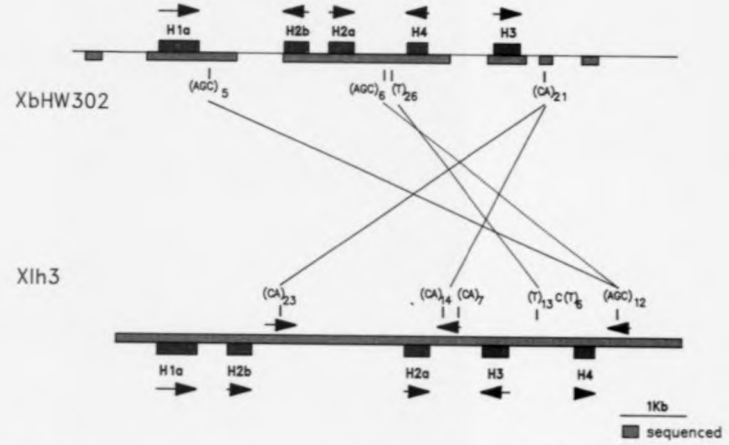


Figure R4 Location of simple repeat sequences common to two histone gene clusters from Xenopus

proposed to be involved with DNA slippage, as a mechanism to explain how they may act as sites for recombination (Hentchel 1982). The alternating purine-pyrimidine repeats may exist in the left-handed Z-DNA conformation, which may also act as foci for recombination events (Nordheim & Rich 1983). Simple repeat elements have been described in association with histone gene clusters in *Xenopus* (Van Dongen *et al* 1984), and other species, (eg newt, see Stephenson 1984). The possible roles and the implications of the occurrence of repetitive DNA in the clusters is discussed in Chapter 6.

These results compliment the gene comparison analyses above, and together they indicate that XbHW302 and X1H3 are related despite displaying contrasting gene arrangements. The interpretation of these results is considered in chapter 6.

3.E MICROINJECTION OF MAJOR CLUSTER CLONES

In order to determine whether the clones contained functional genes, they were microinjected into *Xenopus* oocyte nuclei. The proteins produced were labelled using tritiated Lysine, as described previously (see section 2.B.23). Germinal vesicles were isolated, and the proteins extracted. Two electrophoretic separation systems were employed, namely SDS PAGE and Triton/Acid/Urea (TAU) PAGE. The results are shown in Figures R5 and R6 respectively. In

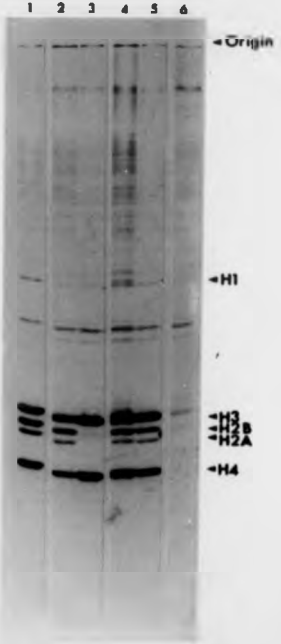


Figure R5 SDS PAGE analysis of ^{3}H lysine labelled proteins after microinjection of the following major cluster clones into *Xenopus* oocytes.

Legend

1	XbHW302	4	XbHW231
2	XbHW321	5	XbHW131
3	XbHW152	6	Uninjected control

each case the tracks display background oocyte translation, plus predominant bands resulting from expression of genes present on the microinjected cloned DNA.

In Figure R5, all the clones except XbHW152 display all the core histones. Clone XbHW152 appears to display no H2B and H2A signals. Clones XbHW302 XbHW131, and XbHW231 appear to produce H1 signals, but as clone XbHW231 displays greater background expression, it is not clear that this H1 signal is a result of expression from the injected clone. The reason for the reduced intensity of H1 and H2A signals is unclear.

Risley and Eckhardt (1981) have identified various *X. laevis* histone H1 subtypes, on the basis of electrophoretic separation. Initial viewing of Figure R5 indicates that clone XbHW302 displays an H1 histone apparent as a fast migrating type. In *X. laevis* H1C was defined as a fast migrating type. As *X. borealis* genes have been microinjected in this study, one can only say that if *X. borealis* H1 subtypes migrate like those of *X. laevis*, then this band could represent the presence of an H1C-like subtype. Analysis of sequence data indicates that clone XbHW302 contains a gene more similar to the *X. laevis* H1A or H1B than the H1C subtype (section 3.D). This supports the belief that mobilities may not be solely determined by the gene sequence. Analysis of the gene sequence may therefore be the better indicator of subtype identity.

Portions of the same samples shown in Figure R5 were also analysed by Triton Acid Urea (TAU) gel electrophoresis. Figures R6a and R6b display respectively short and long exposures of these results. The long exposure was included as the H1 products are not visible on short exposure. This analysis, confirms the presence of all the genes noted above. However, in this case, other minor details emerge. Clones XbMW302 and XbMW231 appear to code for an H2A protein with a greater mobility than those encoded by the other major cluster clones. Secondly, clone XbMW302 appears to show two H1 products. These could result from differences in secondary modification, or could result from expression of two genes. The latter may be less likely, as Southern blot and sequence analysis suggest that only one H1 gene is present on this cluster. Maybe expression of endogenous genes is stimulated by injection of this clone. The appearance of a single H1 band by SDS PAGE indicates a difference between the separation systems.

In this chapter the major histone cluster of *X. borealis* has been characterised in detail by broad analysis of a group of five clones, and fine analysis of a representative member. The comparative analyses with the *X. laevis* clusters and genes, not only set this characterisation in context, but raise questions about the evolution of the *X. borealis* major cluster, which is discussed in Chapter 6. To complement this chapter on the detailed study of a few cloned clusters, chapter 4 considers

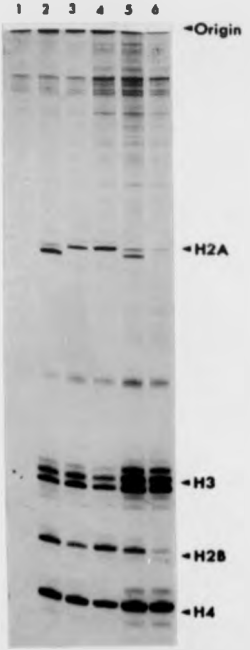


Figure R6a A short exposure of TAU PAGE analysis of ^{3}H lysine labelled proteins after microinjection of the following major cluster clones into *Xenopus* oocytes.

Legend

1	Uninjected control	4	XbMW121
2	XbMW102	5	XbMW231
3	XbMW111	6	XbMW152

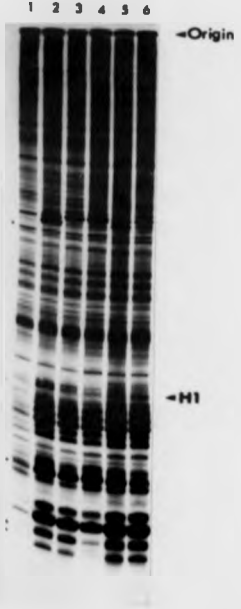


Figure R4b A long exposure of TAU PAGE analysis of ^{3}H lysine labelled proteins after microinjection of the following major cluster clones into *Xenopus* oocytes.

Legend

1	Uninjected control	4	XbHW321
2	XbHW302	5	XbHW211
3	XbHW131	6	XbHW152

the arrangements of this cluster within the *X. borealis* genome.

CHAPTER 4. LARGER SCALE GENOMIC ORGANISATION OF THE MAJOR HISTONE CLUSTER IN *X. BOREALIS*.

4.A PARTIAL RESTRICTION DIGESTS

As previously mentioned, using H4 gene hydridisation the copy number of the histone major cluster in *X. borealis* has already been determined as about 56-63, that is 70% of the 80-90 total H4 gene copies per haploid genome. To address the question of whether these are dispersed throughout the genome or occur in a single tandem array, several approaches were considered. The initial strategy was to try Southern blots on partially SalI digested genomic DNA, extracted from *X. borealis* erythrocytes, using an H4 gene probe. Previous work suggested that the repeated fragment consisted of a 12.8 Kb and a 3.3 Kb component. This was suggested to be the minimum total tandemly repeated unit (Turner & Woodland 1983). Thus if the major cluster were tandemly repeated, a partial SalI digestion probed with an H4 gene, should give a ladder up the autoradiograph. However, despite repeating the experiment many times, no convincing 'ladder' was ever observed, and this strategy was dropped in preference for an attempt at a 'chromosome walk' experiment.

4.B CHROMOSOME WALK

The plan was to obtain a clone bridging two adjacent clusters. To maximise the chance of getting what was wanted, the probe used was from a spacer region at one end of one of the existing clones. It is necessary that this clone is specific, or nearly specific, to the major cluster. Although in theory, a restriction digestion product of an existing lambda clone could be used as a hybridisation probe, in practice spurious hybridisation occurs due to vector contaminants in such probe preparations. This problem can be completely eliminated by subcloning in a different vector.

Figure R2 indicates the position of one M13 subclone SM115, which was originally prepared from an SstI/HindIII double digest of XbHW302, for sequencing purposes, but had not been completely sequenced. It originated from the region flanking the major cluster histone genes, and so was a potential candidate for a chromosome walk experiment. The size of the insert (0.75 Kb) also was within the suitable size range for such an experiment.

To determine the suitability of this subclone, it was tested for the presence of repetitive DNA. Repetitive DNA can bind to elements throughout the genome, and hinder the selective extraction of a clone containing the desired flanking region. Repetitive DNA has been found in close proximity to *Xenopus* histone genes by other groups in this field (Van Dongen *et al* 1984, Perry *et al* 1985), and the

sequencing of the *X. borealis* major cluster revealed several repetitive elements (see section 3.D). Figure R7a shows the hybridisation pattern produced when nick translated SH115 DNA is used to probe a Southern blot carrying restriction digested genomic DNA from *X. borealis*. As the subclone was prepared by a HindIII/SmtI double restriction digestion of clone XbHW302, a HindIII/SmtI double restriction digestion of the genomic DNA was used. Had this DNA fragment contained repetitive DNA, a smear would have been visible, arising from hybridisation to a heterogeneous mixture of fragments from numerous loci throughout the genome. Therefore the presence of a single, discrete band indicates that no repetitive DNA is present in this clone. Thus SH115 is suitable for use as a hybridisation probe in both chromosome walk experiments and Southern blots of genomic DNA.

The size of the hybridisation band in Figure R7a corresponds exactly with the size of the probe. Although this is consistent with the view that this probe is major cluster specific, this appears not be the case. Firstly, when subsequently used as a hybridisation probe on plaque lifts of the library, clones not containing the major cluster were isolated. Secondly, Figure R7b shows a long exposure of the same blot, and it is possible to see above the background non-specific binding, a faint band of 0.67 Kb. The presence of this band indicates that at least some

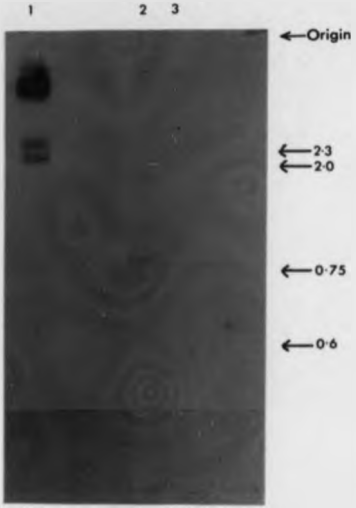


Figure 7a Southern blot analysis of *X.borealis* and *X. laevis* DNA by use of probe SM115
Legend
1 Lambda markers
2 HindIII/BstI digested *X. borealis* DNA
3 HindIII/BstI digested *X. laevis* DNA

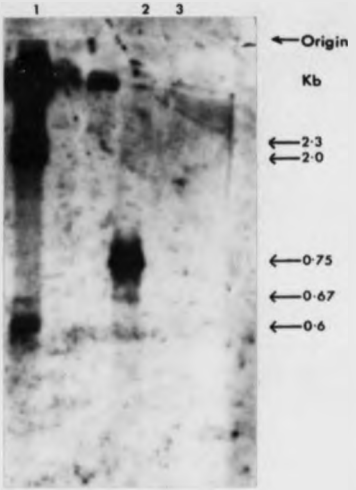


Figure 7b Southern blot analysis of *X.borealis* and *X. laevis* DNA by use of probe SH115. Long exposure of Figure 7a
Legend
1 Lambda markers
2 HindIII/SstI digested *X. borealis* DNA
3 HindIII/SstI digested *X. laevis* DNA

of this 750 bp of sequence, or closely related sequence, is present elsewhere in the *X. borealis* genome, but probably only in one other position.

Track (J) in both figures R7a and R7b contained HindIII/SstI digested *X. laevis* genomic DNA. No hybridisation to *X. laevis* DNA is apparent in either the long and short exposures.

To confirm and extend this observation, DNAs from three *X. borealis* and six *X. laevis* individuals were digested, blotted, and probed with the SH115 fragment. Figure R8 shows the results of this blot experiment. Firstly, the absence of hybridisation to any *X. laevis* DNA confirms that this sequence is *X. borealis* specific. This effect is considered in detail elsewhere (see section 6.D). Secondly, hybridisation to the expected 15 Kb major cluster band is apparent in two of the *X. borealis* tracks. The presence of a doublet in track three, and faint, low molecular weight fragments in tracks 1 and 4 is thought likely to arise from methylation of SalI sites. This is considered in detail elsewhere (section 6.D). Thus SH115 had sufficient specificity for the proposed experiments.

4.C CLONE XbNW102

Having shown subclone SH115 to be a suitable hybridisation probe in a chromosome walk experiment, it remained to screen the lambda 47.1 library with this probe. The library had been plated out in entirety for the first screening, and so

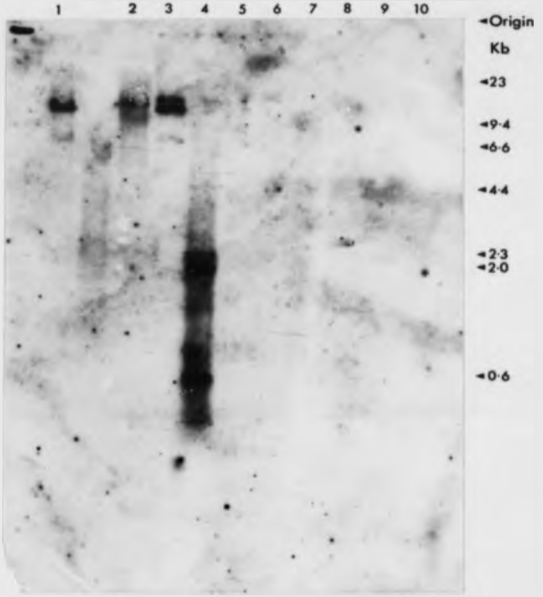


Figure 8a Southern blot analysis of *X.borealis* and *X. laevis* DNA by use of probe SH115.
Legend
1-3 SalI digested *X. borealis* DNA from three individuals
4 Lambda markers
5-10 EcoRI digested *X. laevis* DNA from six individuals

was amplified once. As every positive had been picked on the first screen, the library was potentially histone gene deficient. Despite this, screening of the library with nick translated SH115 did produce positives. Two clones were isolated. Initial restriction digestion of these indicated that one clearly did not contain any bands in common with the major cluster clone XbHW302, and so was not investigated further. The other clone, XbHW102 was mapped for several restriction sites. The restriction site map of clone XbHW102 is shown in Figure R9. The overlap with clone XbHW302 is also illustrated. Clones of inserts in the 8-12 Kb range were isolated from the first screening (Figure R1, R16). Clone XbHW102 fell within this expected size range. Although this clone extended the mapped region of the major cluster further in the desired direction, the extent of this extension was not as great as had been hoped for, and only amounts to some 1.2 Kb.

The isolation of clone XbHW102 was thought to be less likely than the isolation of a clone which extended further in the direction away from the major cluster, for the following reason. Clones carrying H1 and H4 genes had been removed in the initial screen of this library, thus if Sau3A sites existed between the region of SH115 and the H1 gene, one would predict more copies of neighbouring (SH115) clones than clones containing histone genes.

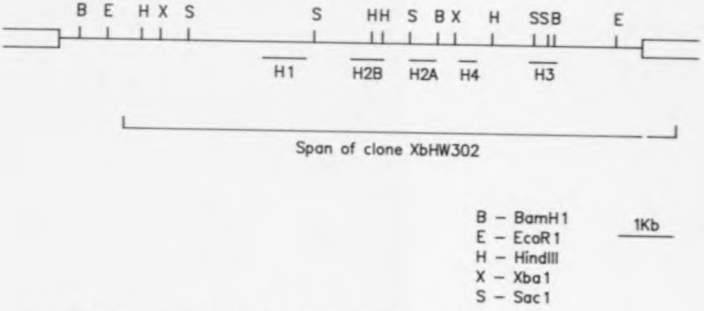


Figure R9 Restriction site map of clone XbHW102

4.D CLONES XbHW9 & XbHW11

Clearly if the major histone clusters were tandemly repeated, with a minimal repeat length of 16 Kb, then chromosome walking using a library with an average insert size of about 9 Kb, would require more clones to be isolated, than if a library with larger inserts were used. This is especially true if clones which only extend small distances, like clone XbHW102, are isolated. Clearly a library of longer average insert length, would reduce the work needed to span any gap in a tandem array. Fortunately, such a library became available (a generous gift from C. Wilson). The insert length was not known accurately, but as a vector capable of carrying large inserts (lambda EMBL 3) was used, and as large 'target' DNA had been selected, it was considered likely to be an improvement. This library was in other respects very similar to the previous lambda 47.1 library, as it contained genomic DNA, partially digested with Sau3A, although a different frog was used to prepare the erythrocyte DNA. Initial screening of this library yielded twelve positives, four of which were purified, and were used to prepare DNA. The sizes of the clones varied between 15 and 18 Kb. Two of them (clones XbHW9 and XbHW11) contained bands which were common to clone XbHW102, and so were mapped in detail. Figure R10 shows the restriction maps of these clones. The figure shows that when the clones are aligned, they represent a region of chromosome of approximately 22Kb.

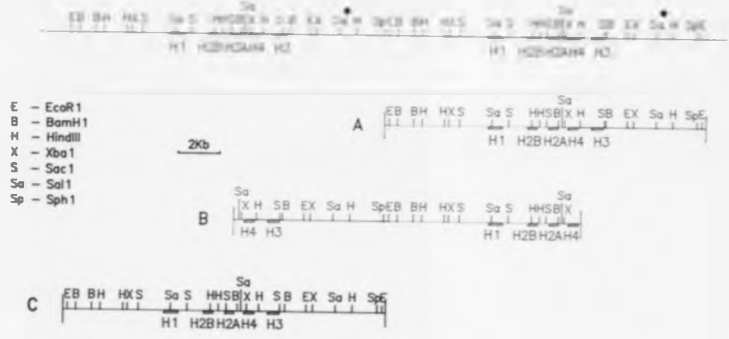


Figure R 10 Restriction site maps of; A Clone XbHW11, B Clone XbHW9
C displays an alternative alignment of XbHW11 relative to XbHW9
The top green line represents a model of cluster arrangements along a chromosome generated by summing the sequences A–C
• indicates a Sal1 site which is cleaved in cloned DNA, but not in genomic DNA (see text)

Clearly the interpretation of the map of clone XbHW9 is that at least in one case in the genome, there are two major histone gene clusters located close to each other, in fact separated by an intergenic spacer region of about 9Kb. Furthermore, the spacer outside of the pair of clusters, is consistent in terms of the restriction map, with another copy of the spacer region occurring between the clusters, suggesting the possibility of another cluster 9 Kb away. The clones can be re-aligned, as indicated in Figure R10.B and R10.C. The top line represents a plausible model of the arrangements of major clusters along the *X. borealis* chromosome, by summing the overlapping alignments of R10.A, R10.B, and R10.C. These results show the repeat length to be 15 Kb.

Comparison of the restriction site maps for clones XbHW9 and XbHW11 with the previously published map (Turner & Woodland 1983), which was based on restriction digests of erythrocyte DNA, both show that for the regions of overlap, almost all the restriction sites occur at identical intervals. This finding was as expected, because probing with histone gene probes on genomic DNA did reveal that the major cluster was very homogeneous ie. discrete bands were observed with each histone probe (Turner & Woodland 1983). One site to not conform is a single SalI site, which appears absent in genomic DNA but present in both cloned clusters. This SalI site is marked * in Figure R10. A fragment of 12.8 Kb is observed in a Southern blot of SalI-digested

erythrocyte genomic DNA, hybridised using H3 or H4 gene probes, while a 4.5 Kb fragment is visible when cloned DNA is used. A possible explanation for this is as follows: SalI is sensitive to methylation at the position marked in the recognition sequence GT(Cm)GAC. *Xenopus* erythrocyte DNA has been shown to be highly methylated (Bird *et al* 1981). Interestingly, it appears as if the genic regions appear un-methylated, while spacer regions are methylated. This finding is consistent with the belief that actively transcribed regions of a chromosome are demethylated, and untranscribed regions are methylated. The strains of *E. coli* used to plate out the library were ones which did not methylate the DNA in this position.

4.E H4 GENE PROBE

The evidence presented above that these clones actually contain the major cluster is only the presence of restriction sites at identical distances to those found in the previously mapped major histone cluster. There is no direct evidence of the presence of histone genes on these clones. It was possible to confirm the map, and thereby the presence of two clusters, by using a nick translated H4 gene as a hybridisation probe, on various digests of both clones. If the maps are correct, part of an H4 gene should be present at either end of clone XbHW9. Thus, an H4 gene used

as a nick translated probe should usually hybridise to two bands in a restriction digest. The sizes of such bands are shown below

Restriction Digest	Clone	Expected Fragment Size
BamH1	XbHW9	2.4Kb, Vector arm +1.3Kb
	XbHW11	2.4Kb
Sal1	XbHW9	4.5Kb, 1.8Kb
	XbHW11	4.5Kb
EcoR1	XbHW9	7.8Kb+Vector arm, 1.5Kb+Vector arm
	XbHW11	9.8Kb

Figure R11 shows the presence of two hybridisation bands for every digest of clone XbHW9, and a single band for clone XbHW11. These bands clearly correspond to the predicted sizes. Tracks 8 - 11 contain DNA from clones XbHW10 and XbHW12. These clones were isolated alongside clones XbHW9 and XbHW11. They each contain an SstI/HindIII fragment of 0.75 Kb, which hybridised to the SM115 probe, but did not contain any other fragments common to the major cluster. From this experiment they appear not to contain any H4 coding DNA.

4.F PAIRS OR TANDEM REPEAT

The characterisation of clones XbHW9 and XbHW11, does not show whether the predicted 55-65 copies of the major cluster occur predominantly singularly, or in pairs scattered

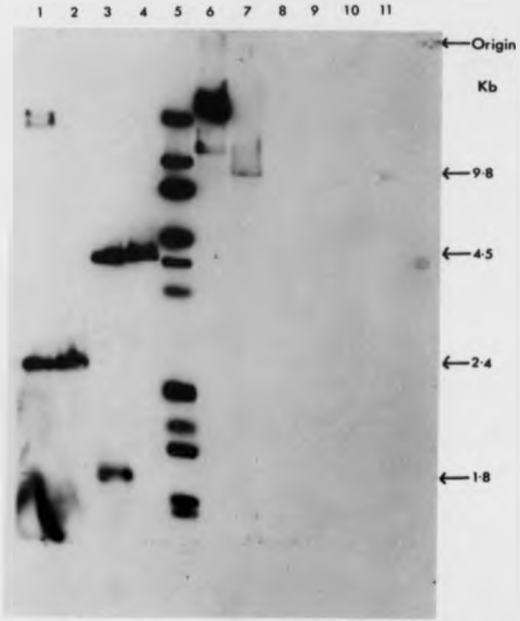


Figure R11 Southern blot analysis of cloned DNA by use of an H4 gene probe.

Legend

Lane		Lane		Lane	
1	XbHW9 BamH1	5	Lambda markers	8	XbHW10 BamH1
2	XbHW11 BamH1	6	XbHW9 EcoR1	9	XbHW12 BamH1
3	XbHW9 Sal1	7	XbHW11 EcoR1	10	XbHW10 HindIII/Sst1
4	XbHW11 Sal1			11	XbHW12 HindIII/Sst1

throughout the genome, or in tandem reiteration. It is simply clear that at least in one instance, a pair of clusters does exist.

To address this problem, it was decided to return to partial digest experiments. With the extended map, and repeat length known (15 Kb), it was possible to attempt to use different restriction enzymes, and correct faults found in the previous attempts. Fresh erythrocyte DNA was prepared, and digested under conditions of varying ratios of restriction enzyme and DNA. Acid treatment of the gel allowed successful transfer of the high molecular weight bands. Figure R12 shows the result of hybridising SH115 insert DNA to a Southern blot of *X. borealis* genomic DNA partially digested with EcoR1.

Interpretation of this figure requires consideration of the predicted products of an EcoR1 partial digestion of a chromosomal region containing tandemly repeated major clusters. This is because the different sized products are expected at different abundances. Figure R13a displays the various sizes of expected products with their derivation. Clearly for the whole repeat length of 15.2 Kb to be cleaved out, requires more restriction sites to be missed (ie 2), than for the formation of species of 11.2 and 13.4 Kb (ie 1). However, the 15.2 Kb band can be produced by cleavage in three ways. This point is illustrated in the figure.

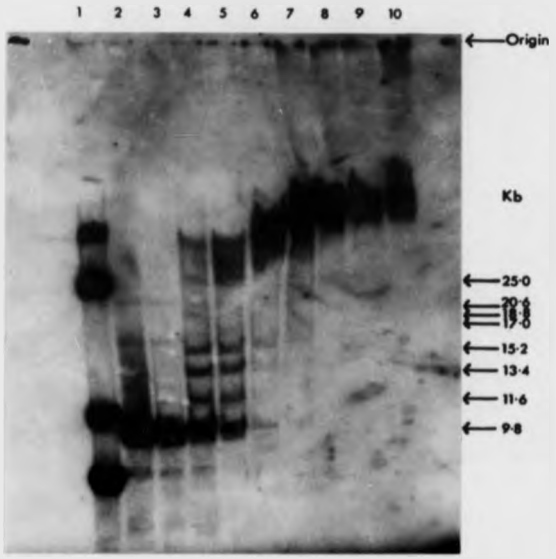


Figure B12 Southern blot analysis of partially EcoR1 digested *X. borealis* DNA by use of probe SM115. Lane values are given in Units of enzyme (U) to ug genomic DNA. The sizes of the products are indicated on the right of the figure.

Legend

Lane		Lane		Lane	
1	Markers	5	0.25 U/ug	9	0.016 U/ug
2	2.0 U/ug	6	0.125 U/ug	10	Undigested
3	1.0 U/ug	7	0.063 U/ug		
4	0.5 U/ug	8	0.032 U/ug		

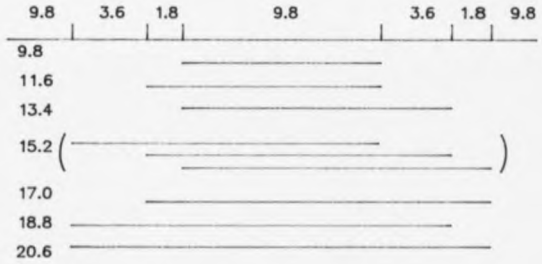


Figure 13.a The expected sizes (Kb) of partial EcoR1 products from digestion of tandemly reiterated major repeats

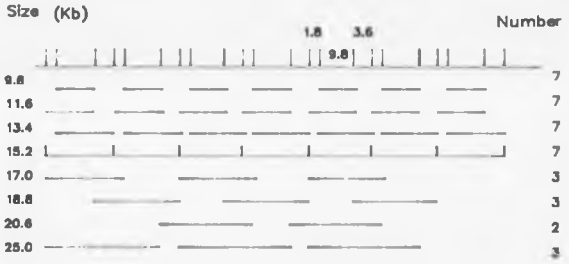


Figure 13.b The theoretical maximum number of partial EcoR1 products from digestion of seven tandemly reiterated major cluster repeats

With this knowledge, Figure R12 can be interpreted. In track 2 (2 U/ug), it is apparent that a total restriction digest displays a single predominant band of 9.8 Kb, the same size as would be predicted from the restriction maps in Figure R10. Measurement of the predominant bands in track 4 and 5 fit the predicted sizes exactly. One can infer from this that the 1.8 Kb EcoR1 fragment and the 3.6 Kb EcoR1 fragment, which represent spacer DNA, and which make up the repeat length as measured from the cloned DNA, do occur in the genome of this frog. They occur in positions flanking the cluster, as predicted from the analysis of the cloned DNA, and they are present in numbers which are sufficient to give the observed hybridisation signal.

Further, very close analysis of tracks 4 & 5, shows faint bands which appear in the size range of the predicted sizes of 17.0, 18.8 and 20.6 Kb. These bands are very faint, and as minor bands are present in the 'total' digest (track 2), little weight should be placed on their occurrence. The occurrence of these bands raises the question as to why they should be so faint relative to the lower bands. By considering the consequences of a partial restriction digestion of a tandemly reiterated 'battery' of major cluster repeats, it is apparent that this is what one would predict. The production of a 17.0 Kb fragment requires sections from two major clusters, hence the number of repeats needed to produce 17.0 Kb fragments is twice that

needed to produce those of 11.2 Kb, or 13.4 Kb. Figure R13b illustrates this point, and displays the maximum number of molecules obtainable from a section of tandemly reiterated 'repeats' containing seven major clusters. Thus, if n represents the total number of tandemly arranged repeats, in a bank, the maximum number of 17.0 Kb fragments obtainable will be approximately n/2 if n is even, and (n/2)+1 if n is odd. There is no suggestion, that this is anything other than a theoretical consideration; the outcome of a partial restriction digest does depend on many factors, including the distance between sites, how exposed the sites are, etc. It would be statistically unlikely, even if an enzyme was cleaving one in five EcoR1 sites, that it would cleave every fifth site. However, this being said, if even in the hypothetical case of maximising the number of obtainable fragments of 17.0 Kb gives a value of half the number of smaller (13.0 Kb) fragments, it can account for the relatively faint intensity of the 17.0, 18.8 and 20.6 Kb bands.

The bands above 20.6 Kb appear stronger. They should contain multiple copies of the Sst1/HindIII 0.75 Kb fragment. One would expect the band of 25.0 Kb to hybridise to twice as much probe per molecule, as any of the smaller ones. This should give a correspondingly stronger signal after autoradiography. However, as the separation is so poor in this region, little weight should be placed on this observation. It is fair to say though, that this result is

not inconsistent with the expected pattern. Despite the long electrophoresis time, the agarose (0.4 %) cannot resolve bands of this size to yield information that can be reliably interpreted. Thus the technique has shown limitations, namely the ability to clearly detect high molecular weight species at low abundance. Agarose gels of below 0.4 % are not only technically difficult to pour, run, and blot, but linearity of mobility against size is progressively lost.

4.G TOTAL GENOMIC RESTRICTION DIGESTS

Clearly there are problems with the partial digest technique, and so other strategies were considered. It has been shown that clusters can be paired, and that the spacer fragment appears to be as abundant as the cluster. However, there is no direct evidence indicating that entire 15 Kb repeats are not scattered throughout the genome. The isolation of clone XbHM9 is the clearest suggestion that repeats may be linked.

It would be possible to construct a cosmid library, with which it could then be shown whether cosmid clones existed with several 15 Kb repeat units linked together. However this again would not give a general picture of the majority of the predicted 55-63 copies of the 15 Kb repeat. However a complete digest of genomic DNA is helpful in resolving this problem. A hybridisation probe from within the cluster, used as a probe on a total digest of genomic DNA, digested by an enzyme which cuts the cluster once, would produce

information on the position of adjacent sites outside that cluster. If the clusters were dispersed throughout the genome, then these sites would be a variety of distances away, and would produce a range of DNA fragments sizes. These would appear as a smear. If, on the other hand, the major clusters were tandemly reiterated, a single fragment of a fixed size would be apparent. From the mapping, one would predict that the fragment size would be 15 Kb.

During the initial mapping no such enzyme had been discovered, and therefore many more six cutter restriction enzymes were tested, to determine whether they could cleave the major cluster at only one site. Figure R14 shows the Southern blot hybridisation pattern of the SH115 probe on *X. borealis* DNA digested with various restriction enzymes. Not all the restriction enzymes have fully digested the DNA. In cases showing enzymes previously used in mapping, the band sizes agree with the mapped distances. SphI appears to produce a band of approximately 15 Kb. Clones XbHW9 and XbHW11 were then mapped for this enzyme, which was found to cut once and therefore be suitable for use in an experiment based on the above rationale. As the reiteration frequency had been described in terms of numbers of H4 genes per genome (Turner & Woodland 1983), and it was possible that some copies of the major cluster lacked the region corresponding to SH115, it was necessary to perform this experiment using an H4 gene probe. This was performed, and Figure R15(a) displays the results of this experiment.

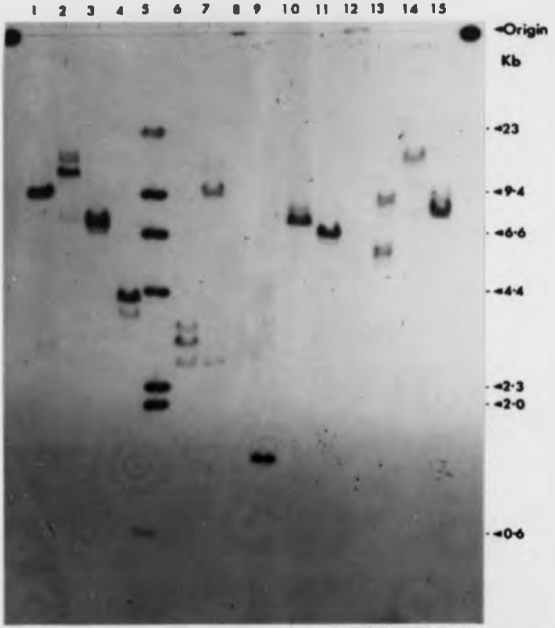


Figure R14 Southern blot analysis of restriction digested *X. borealis* DNA by the use of probe SM115.

Legend

1	EcoR1	4	HindIII	7	Stu1	10	Pvu1	13	Xho1
2	Sal1	5	Markers	8	Tth3.1	11	BglII	14	Sph1
3	BamH1	6	Pst1	9	Ava1	12	Kpn1	15	Sst1

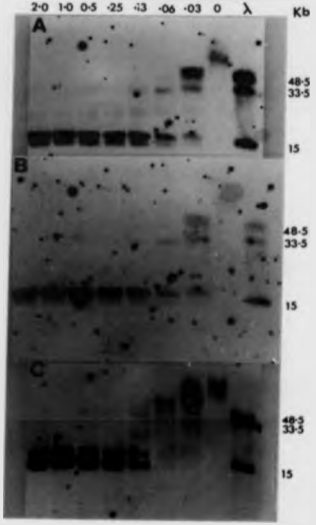


Figure R15 Southern blot analysis of completely and partially SphI digested genomic DNA from *X. borealis* by use of <A> an H4 gene probe, <B> the SH115 probe, and <C> a 'minor cluster' specific probe.

Legend

Tracks are labelled with the enzyme units/ug DNA ratio.

Results

The presence of a predominant band at 15 kb indicates that all the detectable copies of the major cluster are tandemly linked. Further confirmation of this arrangement is indicated by the presence of bands of 30 and 45 Kb in the partial digest tracks. These are consistent with dimers and trimers of the repeat. A band of about 14.2 Kb is also visible in the totally digested sample (left hand) lane. To determine whether this was a major cluster variant, the filter was washed and reprobed with SH115 DNA [R15(b)]. This shows a faint band corresponding to 14.2 Kb, indicating that this lower band could represent a minor variant of the major cluster. However, the relative intensities of upper and lower bands vary between the blots with different probes. For this reason, the experiment was repeated a third time using a minor-cluster specific probe (see chapter 5)[R15(c)]. This revealed a yet smaller band (about 14.0 Kb), which together with the lower band observed in the Southern blot using SH115 probe could account for the lower band in the Southern using an H4 gene probe.

The tandem arrangement of major histone gene clusters is consistent with the results of an *in situ* hybridisation experiment (Turner *et al* 1988). Following hybridisation of an H4 gene probe to mitotic chromosomes of *X. borealis*, a single locus displayed the majority of the labelling.

CHAPTER 5. ANALYSIS OF AN *X. BOREALIS* MINOR HISTONE CLUSTER.

5.A INTRODUCTION

Chapter 3 describes the analysis of one class of histone gene-containing recombinants isolated from a library of *X. borealis* genomic DNA. From the rudimentary restriction maps that were prepared for all the isolated clones, a second set of clones was found, which also shared common fragment sizes. Figure R16 displays the rudimentary restriction maps of these five minor histone cluster clones. The isolation of five clones in this group suggested that this cluster type may be repeated. On this basis it could represent a substantial proportion of the histone genes not present in the major cluster arrangement, and so was chosen for further study.

5.B FINE ANALYSIS OF A MINOR CLUSTER CLONE

The initial analysis of clone XbHW61 by hybridisation indicated the presence of three types of histone gene, namely H1, H3, and H4, all clustered together near the end of the insert (results not shown). The H1 and H3 genes were on a 2.6 Kb HindIII fragment, while the H4 was on the adjacent 3.1 Kb HindIII fragment. Both these fragments were subcloned into phage M13 for sequencing.

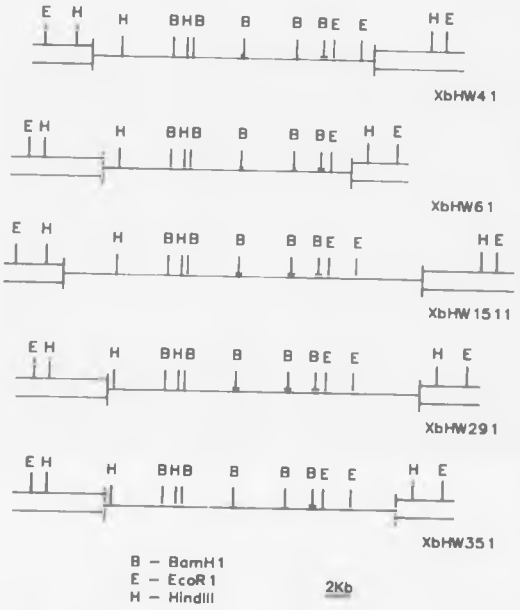


Figure R16 Rudimentary restriction site maps of the minor cluster clones.

The order of fragments bordered by arrows is unknown.

5.C SEQUENCE ANALYSIS OF MINOR CLUSTER CLONE XbHW61

The minor cluster was sequenced for several reasons. Firstly to compliment and extend the study of histone genes in *X. borealis*. Secondly the H1 gene of clone XbHW61 displayed several expression products with different electrophoretic migration, on microinjection into the *Xenopus* oocyte (see Section 5.D). Sequence analysis could reveal any unusual features, either in the coding, or flanking regions, to account for the observed expression pattern. Thirdly, Woodland *et al* (1984) have determined the sequence of a minor H4 message class, found in all tissues (see Section 1.B.8c). Sequence analysis of the minor gene cluster in the upstream H4 gene region, would indicate if the observed expression was from this cluster type. Sequencing was originally performed following a shotgun cloning procedure, but selective cloning of small fragments made this strategy inefficient. A second strategy of cloning known restriction fragments was chosen in preference. Appendix A5 displays the sequence of the genic region of clone XbHW61, comprising some 3850 base pairs. To verify the sequence derived from one strand, almost all of the complimentary strand was sequenced as well, as indicated in figure 17. Each gel run was merged using the Microgenie software package. In short, the sequence confirmed the existence and location of three histone genes, an H4, an H3 and an H1 gene. The arrangement of these genes is indicated in figure 17.

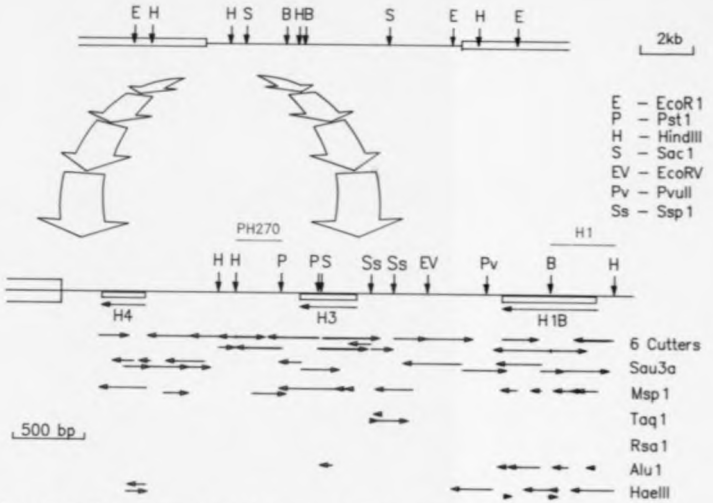


Figure R17 Map of clone XbHW61 displaying location and orientation of the sequenced M13 subclones

Sequence analysis;

Genic Regions

H1 gene

X. laevis H1 subtypes were originally designated as types A-C on the basis of different electrophoretic migration (Rimely & Eckhardt 1981). Thus any subtype designation of *X. borealis* H1 histones must be consistent with the *X. laevis* subtypes, and must imply relatedness rather than similar elctrophorectic behaviour. The identity of the XbHW61.H1 subtype was determined by comparison with published *X. laevis* H1 subtype sequences. Appendix A7(a) to (f) displays alignment of the H1 coding sequence from XbHW61 with sequences of the *X. laevis* H1A, H1B (Perry *et al* 1985), and H1C (Turner *et al* 1983) subtypes. This figure in the appendix also contains comparisons of the predicted translation products in every case. The DNA sequence match/length alignments of H1B(93%), H1A(72%) and H1C(83%) clearly indicate that clone XbHW61 contains an H1 subtype equivalent to the *X. laevis* H1B. Overall the XbHW61.H1B gene product is one amino acid residue shorter than the Xlh1.H1B, and a total of fifteen bases can not be aligned between the sequences, suggesting an insertion or loss of five codons since divergence. It is possible that this difference could account for the production of three H1 products on microinjection of this clone, as seen in figure 21b, but no obvious mechanism is apparent.

H3 and H4 genes

The H4 and H3 genes display typical histone gene features. Appendix A6(b)(c) displays the predicted translation products.

Flanking sequences

The H3 gene displays the sequence CCATTCT at -39, which is similar to the consensus PyCATTCPu 'Cap' element. The H1(b) and H4 genes display GAGTTT (-41bp) and CAGTTT (-30bp) respectively where one would expect the 'Cap' motif. These latter two sequences appear to be related to the GTGTTT found in the corresponding region of XbHW302.H1 (see section 3.D). The XbHW61 histone genes display the TATAA motif, at -59bp(H3 and H4 genes) and -65bp(H1 gene), while consensus CCAAT elements can be found at -81bp(H3 gene), -91bp(H1 gene), and -98bp(H4 gene). These features, like the 3' motifs for each gene, are typical of functional histone genes, and there is no obvious sequence feature that could account for multiple expression products noted after microinjection of these clones (see section 5D). It seems most likely that post translational modification products account for the observed band pattern.

Analysis of H4 mRNA leader sequences revealed a minor class, which displayed two sequence forms; A and B (see 1.B.8c, Woodland et al 1984). The sequence of these two forms is shown below together with the corresponding XbHW61 sequence.

CGCTCg/tCTTGC T ATC ATG - A type

C G CTTTAg/tGTC ATG - B type

ACATC T CTTGC T ATC ATG - XbHW61

From this it is clear that XbHW61 codes for an A-like type mRNA, although interestingly, differences are apparent at the 5' end. The primer extension sequence represents a consensus from many RNA molecules, which could result from expression from a variety of genes displaying variable leader sequences. Genes that had diverged to the extent that they displayed leaders of different lengths would be unlikely to co-migrate on the gel. Obtaining the mRNA leader sequence data was difficult because the primer extended products were at trace levels, and were consequently very faint and hard to read (M. R. Woodland pers. comm.). Bearing these points in mind, it appears likely that XbHW61 is an example of one of the gene types which was expressed to produce the A type consensus, although the majority of molecules displayed a different consensus at three positions at the 5' end. The alternative argument, namely that the sequence differences exclude XbHW61 from this A-type class, demands that this cluster type is not expressed as the predominant message of this leader length, in any of the tissues (including ovary) or stages studied. The microinjection into oocytes (ie ovary tissue), confirms *in vivo* expression (section 5.D) of XbHW61, albeit at high copy number. That the A type was found in all

tissues and developmental stages studied, suggests a ubiquitous expression profile for the XbHW61 cluster type, without tissue-specific or developmental regulation. Thus XbHW61 and the major cluster clones (Chapter 3) appear to have similar expression profiles.

Sequence comparison

Of more interest than the sequence itself, are the comparisons with other sequences. The first comparison was between the H4 genes and between the H3 genes of the following four cloned clusters: *X. laevis*; Xlh1 and Xlh3 (Perry *et al* 1985) and *X. borealis*; XbHW302 and XbHW61. The aim of these analyses was to shed light on the recent evolution of these clusters. The H3 and H4 genes were chosen in preference of H1 genes, as any recent changes could be masked by the differences in H1 subtypes. The degree of similarity was determined by pair-wise alignment of the sequences. The complete H4 comparisons are displayed in the appendix figure A8 to illustrate this pair-wise alignment. A summary of the data is given in figure R18. Figure R18 also shows corrected values for the percentage divergences. The reason for needing to correct these values is as follows. As the sequences of two genes diverge, the probability of any base position being mutated more than once increases. Consequently, a direct comparison of divergence will under-estimate the number of changes that have occurred. The extent of this under-estimation is non-linear, such that the

H4

CLONES	XbHW61	XbHW302	Xlh1	Xlh3	
XbHW61	–	51.98	19.93	59.96	Corrected
XbHW302	14.6	–	53.47	24.25	
Xlh1	8.4	12.6	–	50.87	
Xlh3	15.5	8.1	13.3	–	

Uncorrected

A

H3

XbHW61	–	62.4	17.9	43.5	Corrected
XbHW302	14.9	–	64.9	27.0	
Xlh1	5.6	15.4	–	43.0	
Xlh3	11.8	8.0	11.8	–	

Uncorrected

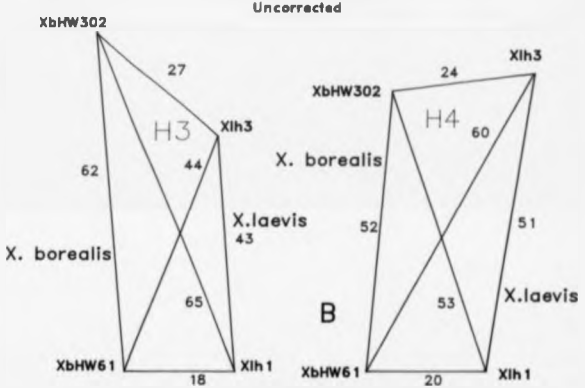


Figure 18 A: Tables displaying the percentage divergence between the silent sites of the coding regions of Xenopus H3 and H4 histone genes by separate alignment of each sequence with each other sequence. The values in red are corrected for multiple changes in these sites.

B: Spacial diagram based on, and displaying the corrected figures from part A. Part B also indicates the species from which each cloned cluster originated. Details in text.

more dissimilar two sequences appear, the greater the under-estimation. The amino acid sequences of the H4 and H3 are conserved absolutely in these clusters, so all nucleotide divergence occurs in the silent sites. In H4, for example, these total some 26% of the DNA sequence. Thus the corrected value refers to the percentage of silent site changes summed over the total potential silent sites, corrected for multiple hits (see 2.B.27).

From the analysis of H4 genes several points emerge:

1> Gene XbMW61.H4 and X1h1.H4 appear more closely related to each other than either does to the other cloned H4 genes within that same species (ie XbMW302.H4 and X1h3.H4 respectively).

2> Interestingly, each of these cluster pairs (XbMW61/X1h1 and XbMW302/X1h3) carry the same H1 gene subtype. This suggests a greater homology between cluster types, as based on H1 subtype, than between clusters within the species.

3> The apparently closely related XbMW61.H4 and X1h1.H4 genes are present on clusters which display a similar topology (gene order, polarity and intergenic distance). This is in contrast to XbMW302.H4 and X1h3.H4, which although only slightly less similar in nucleotide sequence, display completely different cluster structures. All these points hold true for the H3 gene analyses as well, suggesting that the cluster environment affects the evolution of each gene. Interestingly, the degree of

similarity differs between the corresponding H3 and H4 gene pair-wise comparisons. This also is not wholly unexpected, as the rate of silent site divergence of a gene is unlikely to be dependant on the cluster environment alone. The inferences of this finding are considered in Chapter 6.

The second comparison was of the entire corresponding genic regions of clusters XbHW61 (3850 bp) and Xlh1 (4255 bp). The results of this matrix plot are shown in figure R19. Several interesting points emerge from this analysis. Firstly, the regions corresponding to the genes appear almost on a straight line, indicating an approximate similar intergenic distance in the two clusters. Two small deletions have occurred in XbHW61 relative to Xlh1 (or insertion *vice versa*). One between the H1 and H3 genes is about 180 bps, while the second is smaller (about 40 bps) and is seen between the H3 and H4 genes. Secondly the intergenic sequences display stretches of high and low homology within each region. The high homology regions extend 200-400 bp upstream from the genes. Clearly this contrasts with the corresponding spacer sequence comparisons between XbHW302 and Xlh3. The implications of these results are considered in the discussion (see section 6.C).

5.D MICROINJECTION OF MINOR CLUSTER CLONES

To determine whether the minor cluster clones were functional they were microinjected into *Xenopus* oocytes. The resulting proteins were analysed in a fashion identical to

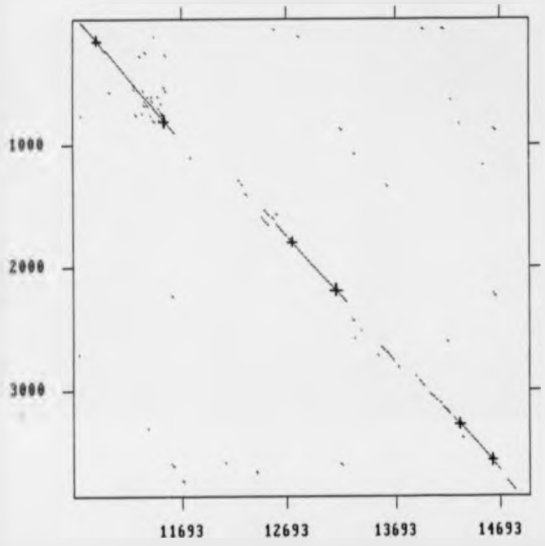


Figure R19: Matrix homology plot of XbMW61 [ordinate] against XlhI [abscissa] sequences. + marks the coding regions. The parameters were set to search for an 80% match over 15 base pairs.

the major cluster clones (see section 3.E). The clones in the minor cluster family are: XbHW41, XbHW61, XbHW1511, XbHW291, and XbHW351 (Figure R16).
Unlike the others in this group, clones XbHW41 and XbHW1511 proved difficult to prepare, and so were not analysed by this method.

The results of this SDS PAGE analysis are displayed in Figure 20. Separation of the clone XbHW61 products shows the presence of H3, H4, and what appear to be two H1 products (although this is clearer on the autoradiograph). These products appear as fast and slow migrating subtypes. XbHW291 displays a faint H1 product, but as the all other proteins in this track appear fainter than those of the other tracks, the H1 band probably does indicate stimulated expression. The histone H1 from XbHW351 is particularly fast migrating which may indicate the presence of an analogue of the *X. laevis* H1C subtype. Like clone XbHW61, both clones XbHW291 and XbHW351 display H3 products. However unlike clone XbHW61, neither displays strong H4 expression. This expression pattern can be explained by considering the restriction map of clone XBHW61, in conjunction with the rudimentary restriction maps of the other clones (figure R16), which show that the 5' H4 gene sequences are likely to be absent in XbHW291 and XbHW351.

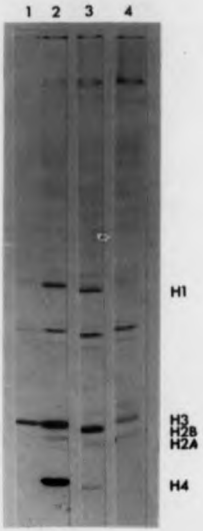


Figure R20 SDS PAGE analysis of ^{3}H lysine labelled proteins after microinjection of the following minor cluster clones into *Xenopus* oocytes.

Legend

1	XbHW291	3	XbHW351
2	XbHW61	4	Uninjected control

The results of TAU PAGE are displayed in Figures 21a and 21b. Figure 21b was included because the H1 signal appeared toofaint on the short exposure. TAU PAGE of these minor cluster clones again confirms the presence of histone types detected by SDS PAGE. Two points of interest also emerge. Firstly XbHW61 appears to produce an highly mobile H2A, similar to that of the major cluster clone XbHW302 (see Figure R6a). As Southern blot hybridisation with an H2A probe did not reveal any H2A sequence (results not shown), it is difficult to account for the presence of this band. Possibly this clone contains a gene coding for a protein similar to the H2Af isolated from chicken (Harvey *et al* 1983). Southern blot analysis only allows for the detection of sequences which are highly homologous to the hybridisation probe. An H2Af-like gene may not hybridise to the *X. laevis* H2A gene which was used in these experiments. Secondly, clone XbHW61 appears to also synthesise both a fast and a slow migrating H1 protein. The presence of the two H1 products on both separation systems prompted further investigation on clone XbHW61 to determine if this effect was caused by two separate genes. Clone XbHW61 has already been shown from sequence analysis to contain a gene that has strong homology to the *X. laevis* H1B type. However, the analysis of one gene

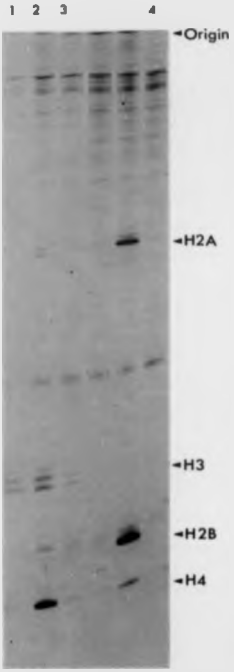


Figure R11a Short exposure of TAU PAGE analysis of ^{3}H lysine labelled proteins after microinjection of the following minor cluster clones into *Xenopus* oocytes.

Legend

1	XbMW291	3	XbMW351
2	XbMW61	4	Uninjected control

Clones not fitting the minor cluster pattern were included in the unlabelled tracks.

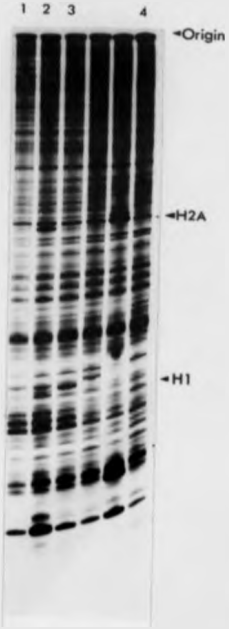


Figure R21b Long exposure of TAU PAGE analysis of ^{3}H lysine labelled proteins after microinjection of the following minor cluster clones into *Xenopus* oocytes.

Legend

1	XbHW291	3	XbHW351
2	XbHW61	4	Uninjected control

Clones not fitting the minor cluster pattern were included in the unlabelled tracks.

does not rule out the possibility of a second gene on this cluster.

5.E H1 GENES

The question of whether two histone H1 genes existed within this clone remained. The formation of the restriction map, and characterisation of the region coding for the H1 histone allowed this question to be addressed. An H1 gene probe was prepared from the HindIII/BamH1 483 bp fragment (Figure R19). This fragment contained almost all the conserved region. It was used as a probe on a Southern blot at low stringency, to allow for cross hybridisation to any different histone H1 isotype that might be present elsewhere on clone XbHW61. The results of this experiment are shown in Figure R22. For two H1 genes to occur and give a single hybridisation band under these conditions requires that both genes be present on a single fragment, or that two fragments migrate together. The single bands in tracks 3 and 5 are so small that they could not carry two genes, and would have to be doublets. Restriction map data rules out this possibility. One must conclude that only a single H1 gene is present on this clone. Thus the most likely reason for the appearance of two histone protein products from the SDS and TAU PAGE analysis is secondary modification.

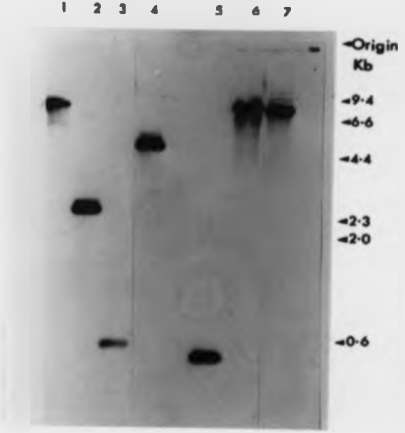


Figure R22 Southern blot analysis of XbHW61 DNA digested with the following enzymes, probed with an H1 gene probe.

Legend

1	EcoR1	4	Sst1	7	Sph1
2	HindIII	5	HindIII/BamH1		
3	BamH1	6	BglII		

5.F ORGANISATION OF THE MINOR CLUSTER WITHIN THE GENOME

Since five apparently closely related minor cluster clones were isolated, it seemed likely that this cluster type may be repeated several fold in the *X. borealis* genome. This prompted further investigation. To this end, a probe (PH270) was prepared from the M13 2.5 Kb HindIII subclone insert. This probe was a PstI/HindIII fragment, corresponding to the downstream region from of the H3 gene (see Figure R17). Southern blot analysis was utilised to test whether this sequence was present in the major cluster, and hence useless for this purpose. No cross reactivity was found with the *X. borealis* major cluster clones, on short exposure, however, faint bands were visible on long exposure. From the sequence comparison (figure R19), it was known that this region displayed weak homology with the corresponding region of the *X.laevis* clone Xlh1. In conclusion, the PH270 hybridisation probe is sufficiently specific to the minor clone to only allow detection of other copies of this cluster in the *Xenopus* genome, by short exposure. Figure R23 shows the results of using probe PH270 in Southern blot analysis of restricted *X. borealis* genomic DNA. The main point of interest from this result is the presence of distinct single bands in various lanes. The sizes of these bands appear specific to this cluster type. Neither the *X. laevis* clones Xlh1, and Xlh3 (Perry *et al* 1985), nor any of the seven *X. laevis* clones isolated by Old *et al* (1982) have fragments of these sizes.

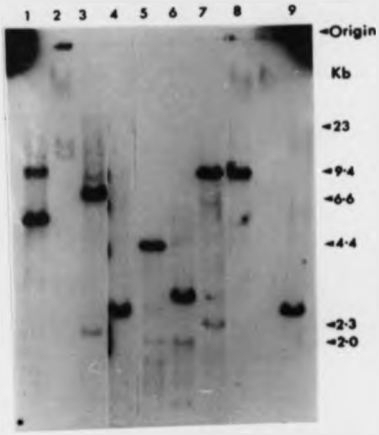


Figure R11 Southern blot analysis of *X. borealis* DNA digested with the following enzymes, using the PH270 probe

Legend

1	EcoR1	4	PstI	7	BglII
2	SalI	5	AvaII	8	XhoI
3	BamHI	6	PvuII	9	SstI

From the matrix sequence comparisons, it was known that the PH270 probe region is not highly similar to the equivalent region of the Xlh1 cluster, yet the two clusters do appear similar. By using this probe on *X. laevis* DNA, it was possible to ask whether this sequence occurred elsewhere in the *X. laevis* genome. The results of this experiment using the minor cluster probe are displayed in Figure R24. The restriction enzyme EcoR1 was chosen in this case, as the results in Figure R23 had indicated that a strong signal from *X. borealis* could be obtained with this enzyme. Clearly no highly homologous sequence element is present in *X. laevis*. Two further points of interest emerge from this analysis. Firstly, it is clear that faint hybridisation does occur in *X. laevis*. This is homogeneous between all six individuals tested and is 9.2 Kb in length. However, this hybridisation is clearly considerably weaker than in *X. borealis*, and was only apparent on a long exposure.

The second point of interest is that the different *X. borealis* individuals appear to display heterogeneity in the presence of an EcoR1 site within this sequence. Another possibility is that two minor types of cluster exist which display part or all of this sequence. Frog 1 could represent a homozygote, while frogs 2 and 3 could be heterozygotes. Neither of the bands in tracks 2 and 3 are of the same size as the faint *X. laevis* band.

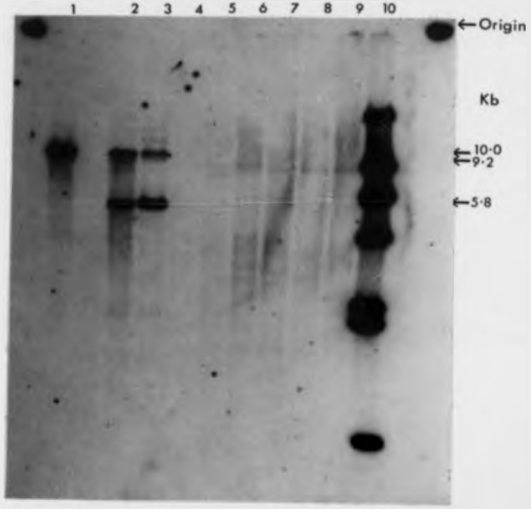


Figure R24 Southern blot analysis of EcoR1 digested genomic *X.borealis* and *X. laevis* DNA by use of probe PH270

Legend

1-3	*X. borealis* DNA from three individuals
4-9	*X. laevis* DNA from six individuals
10	Lambda markers

To conclude, the analyses presented in this chapter complement and extend those of Chapter 3, to cover *X. borealis* histone genes not in the major cluster arrangement. Although more is known about the chromosomal organisation of the major cluster (Chapter 4), the minor cluster was analysed at gene and cluster level. From extensive comparative analyses, this minor cluster appears to share more common features with the *X. laevis* cloned cluster Xlh1, than the *X. borealis* major cluster, which in turn, displays a certain similarities to the *X. laevis* cloned cluster Xlh3. Chapter 6 discusses these findings.

CHAPTER 6. DISCUSSION

6.A. INTRODUCTION TO DISCUSSION

The results chronicle an investigation into the predominant histone gene clusters of *X. borealis*. The isolation and analysis of various clones containing histone genes from *X. borealis* is described. These clones were allocated to the classes of 'major', and 'minor' cluster clones.

Major cluster clones were restriction site mapped, and both genic and intergenic regions were partially sequenced. The sequences showed typical histone gene and spacer features, and microinjection analysis confirmed these genes to be functional. Comparisons of genic and intergenic sequences revealed homology to *X. laevis* cluster Xlh3. Further clones isolated by chromosome walking indicated that the major clusters were closely packed in a tandemly repeated chromosomal arrangement, and this was confirmed by Southern analysis of genomic DNA. Further Southern analysis using a cluster-specific element failed to hybridise to any *X. laevis* sequence, indicating that this cluster was at least in totality *X. borealis* specific.

Minor cluster clones were similarly mapped and the entire genic region of clone XbHW61 was sequenced. The presence and identity of a single H1 gene was determined in this clone, and an H4 and H3 gene were characterised. Comparison of a cluster-specific clone again indicated no detailed intergenic homology with *X. laevis*, but an overall

similarity to an *X. laevis* clone was observed. In this section, the significance of these results will be discussed in the context of related work.

6.B INTRODUCTION TO THE EVOLUTION OF HISTONE GENE FAMILIES

One of the consequences of the extensive investigations which have elucidated histone gene structure in such a wide range of genera, is that evolutionary pathways and mechanisms can be studied. The initial finding that the histone gene clusters of *X. borealis* differed so markedly from *X. laevis*. (Turner & Woodland 1983), was the reason for this detailed investigation. Thus it is in the light of the histone gene organisation of *X. laevis* that the results in this thesis will be considered, and it is why *X. laevis* sequences were used in the comparisons in chapters 3 and 5. Section 1.B.8 of the introduction reviews the organisation of histones in *X. laevis*. The key points here relate to the exact differences between *X. laevis* and *X. borealis* histone gene clusters, and how these can be accounted for in terms of evolution. Clearly one has to account for the apparent rapidity and extent of change. The factors shaping *Xenopus* histone gene evolution could have played a role in the evolution of histone genes across the wide range of species studied. Naturally one also seeks to determine what molecular mechanisms are involved.

What is the functional significance of these differences? These two species are so closely related that they can successfully interbreed. Clearly the phenotypic consequences of the different patterns of histone gene organisation must not be so fundamentally significant as to affect viability seriously. This begs the question of the importance of different cluster structures. There is little evidence as to whether there is the functional role of different histone cluster structures in *Xenopus*. Although there appears to be a link between H1 subtype and cluster structure in *X. laevis* (Destree *et al* 1984, Perry *et al* 1985), it is not apparent that these arrangements have functional significance. Woodland *et al* (1984) have addressed this problem by asking whether expression of the H4 gene class changes through development, or between tissues (see section 1.B.8). In chapters 3 and 5 it was shown that the ubiquitous message classes described by Woodland *et al* are coded for by the major and minor cluster types in *X. borealis*. It seems that cluster structure differences do not affect expression profiles.

Zernik *et al* (1980) suggested that invertebrates displayed highly conserved histone gene cluster arrangements, while those of vertebrates were highly diverged. They also proposed that this arose due to the effects of different selection pressures. The finding that the major *X. borealis* cluster is tandemly repeated does not support this theory, nor does the discovery and

characterisation of the sea urchin late genes, which display gross heterogeneity (see section 1.B.3). Furthermore, *H. viridescens* has a highly homogeneous cluster structure (Stephenson *et al* 1981a).

Stephenson *et al* (1981b) questioned whether homogeneity between different histone gene clusters was linked to high copy number. This too seems unlikely. Firstly, *X. laevis* and *X. borealis* both display approximately the same number of histone genes, but clearly differ in homogeneity. Secondly, this number is only slightly less than that quoted for *Drosophila*, which has highly homogeneous repeated histone gene clusters.

Inevitably it appears that the evolutionary behaviour of single or low copy sequences is easier to follow than multicopy and multigene sequence elements. The histone genes of several species fit into the latter category. In *Drosophila*, Steinmann (1982) showed that histone genes, and other multicopy elements displayed greater mobility in chromosomal location between closely related species than single or low copy sequence elements (see section 1.B.4). Secondly, work by Coen *et al* (1982) also revealed that histone genes and other repeated elements evolved rapidly.

In *Xenopus*, the analysis of *X. laevis* genomic DNA from six individuals by Turner *et al* (1983) displayed the wide heterogeneity of histone cluster structures. This finding has been borne out by the variation in cluster

structure of cloned isolates (see section 1.B.8). The wide heterogeneity both within and between individuals clearly points toward rapid evolutionary processes, which translocate and duplicate genes around rapidly into new cluster conformations in *X. laevis*. In contrast, *X. borealis* displays a highly homogeneous arrangement in the form of the major cluster, that carries the majority of the histone genes, in every individual studied (see chapter 4). However, the gene arrangement of the *X. borealis* major cluster is unlike any other published, so probably arose recently, presumably by the same forces which generate the diversity observed in *X. laevis*. Thus it would then appear that in *Xenopus* variant clusters arise by a set of gene shuffling processes. One such variant was then multiplied up, and spread to form the major cluster in *X. borealis*, by a second set of processes. Both these sets of processes will be considered, in sections, 6.C., and 6.D. respectively.

6.C. GENERATION OF CLUSTER VARIANTS

The processes that rapidly generate new cluster conformations can be studied by detailed analysis of clusters. The analyses in chapters 3 and 5 reveal the similarity of clusters, and hence presumably how related they are.

Clearly there are limitations to inferences made on the comparison of only two histone genes clusters from *X. laevis*, when a wide range of cluster structures are known to

exist. If the range and frequency of all different cluster structures in *X. laevis*, was completely characterised, it might be possible to order all the intermediates in the gene shuffling processes. One might then be able to track all the recent histone gene movements and elucidate mechanisms. However, one is confined to considering the end products of evolutionary processes. I have limited this study to comparisons between those *X. laevis* clusters for which the most complete sequence data is available. As the evolution of gene sequences is generally more constrained than spacer sequences, each of these classes of DNA will be considered separately.

6.C.1 COMPARISONS BETWEEN *XENOPUS* HISTONE GENE CODING SEQUENCES

XbHW61/X1h1 CLUSTERS

The XbHW61/X1h1 comparison is considered first. Firstly, the clusters display the same overall topology; they share gene order, polarity, and have similar intergenic distances. Secondly, analysis with other H4 genes show that XbHW61.h4 is more similar to X1h1.H4, in terms of silent site divergence, than any other pair of H4 genes analysed in this study. This was found to be the case for the H3 genes as well. However, the complete series of H4 and H3 gene comparisons also revealed that cluster organisation seems unrelated to how diverged the component genes are; XbHW302

and Xlh3 show different cluster topologies, but closely related H4 and H3 genes. This is an expected consequence of the histone gene shuffling process in *Xenopus*.

The silent site divergences observed between the H3 and H4 genes of XbHW61 and Xlh1, are 18% and 20% respectively. These genes could have diverged from a common ancestor at the same time and have evolved at the slightly different rates, or alternatively they could have diverged at different times, and evolved at the same rate. There are several reasons why the former is more likely. Firstly, as the genes are so close, it is likely that when one is duplicated, the other would be as well. Secondly, there is no reason to think that H4 and H3 genes should evolve at exactly the same rate, although one might expect them to evolve at similar rates. Recent work has shown the rate of silent substitution to correlate with codon usage (Ticher & Graur 1989), in certain rat and human genes. Codon usage analysis for the XbHW61 genes, displayed different biases for the H4 and H3 genes (results not shown). The recent finding that part of a murine H3 coding sequence was required for high level expression (Hurt *et al* 1989), can explain the importance of codon choice, and why silent site divergence rates can vary. The divergence rates are sufficiently similar to suggest that gene conversion or correction has not happened to either one of these genes. Consequently the difference in the degree of silent site divergence between the histone gene types in the XbHW61/Xlh1

comparison, should not be taken to suggest that the genes present on the clusters today, have not all evolved together from a progenitor cluster.

XbHW302/X1h3 CLUSTERS

At the cluster level, the XbHW302/X1h3 comparison contrasts with the XbHW61/X1h1 comparison; gene order polarity and intergenic distance suggest that XbHW302 and X1h3 are completely unrelated. However, nucleotide analysis reveals a more complete picture. Figure R18 shows that the H4 and H3 genes of XbHW302 and X1h3 are very similar. This prompted the comparisons between all the histone genes on these clusters (see section 3D). For each histone gene type, a higher match/length ratio was observed between XbHW302 and X1h3 than with X1h1. Thus, although it is possible that gene exchange between other cluster types occurred, the consistently high degree of similarity, relative to comparison with either XbHW61 or X1h1, for all five genes, suggests that the genes on XbHW302 and X1h3 diverged from common ancestors, and that these clusters could be related.

6.C.2 COMPARISONS BETWEEN *XENOPUS* HISTONE GENE SPACER SEQUENCES

Spacer elements are known to be less constrained to diverge than coding regions and consequently diverge faster. Hence high spacer sequence homologies provide stronger evidence for two sequences having recently shared a common ancestral sequence.

XbHW61/Xlh1

The intergenic regions of XbHW61 and Xlh1 share substantial regions of homology. Comparisons between unrelated or distantly related clusters show only specific elements. Testament to this in *Xenopus* is the comparison by Perry *et al* of Xlh1 and Xlh3, which shows the conservation of specific elements of only a few bases in length. The occurrence of substantial, well conserved intergenic sequences in the Xlh1/XbHW61 comparisons supports the argument for Xlh1 and XbHW61 to have recently arisen by divergence from an ancestral cluster carrying ancestral copies of the genes presently found on these clusters. This being so, one would expect the high level of homology to extend throughout the intergenic regions. It does not. In the central regions between the genes there appear to be completely unrelated DNA sequences (figure R19). Even if the earliest estimated time since divergence of *X. laevis* and *X. borealis*, which is 20 Myr (see 1.A.2), and a high nucleotide divergence rate of 1%/Myr is taken, one would expect 80% similarity, although this is assuming that the clusters are

not paralogous. This is an interesting finding when considered in the light of the behaviour of other elements. Comparisons between *X. laevis* and *X. borealis* spacer sequences upstream of a single or very low copy actin skeletal gene display high sequence conservation as far as has been sequenced (about 500bp)(M. Boardman pers. comm.). All this intergenic DNA appears to evolve at a constant rate. In *X. laevis*, the copy number of Xlh1 has been determined to be about 10 copies per haploid genome (Perry *et al* 1985). Although the copy number of XbMW61 is not known in *X. borealis*, the isolation of five clones, together with the Southern blot result in figure R15, suggest that it is repeated several fold. Thus clearly there is a difference between the single or low copy actin gene far upstream sequence evolution, and that of the moderately repeated minor histone gene cluster type. Although, it must be borne in mind that several cytogenetic processes (eg gene conversion and gene correction) may shape the evolution of multicopy sequence, but not single copy sequences.

At the other extreme are the very highly repeated 5S RNA genes (see Fedoroff 1979). In *X. laevis* there appear to be about 24,000 copies, while in *X. borealis*, about 9000. Comparisons between these elements show completely diverged spacer sequences. The minor clusters XbMW61 and Xlh1 appear to be intermediate. Thus the rate of spacer divergence could be a function of copy number, and it would appear that when sequences multiply to several fold in genomes, the first

sequence changes are found distant from the genes. As a sequence element evolves to be high copy number, the majority of non-transcribed regions diverge. The function of much intergenic DNA is poorly understood, but more functional elements are continually being uncovered in the region lying up to a few hundred bases 5' to the gene (see section 1.D.2). The finding that the distal upstream sequence appears to have diverged while those closer have not suggests that these far upstream (or distal downstream) regions are freer to diverge, ie the conservation is maintained by selection.

Comparisons between intergenic spacer of XbHW61 and XbHW302 revealed no strong homology of over 30 bps in length; presumably these clusters are unrelated. The low degree of coding sequence similarity supports this.

XbHW302/Xlh3 CLUSTERS

The gene sequence comparisons suggest that these two clusters could have once shared a common ancestor. This view is strongly supported by the XbHW302/Xlh3 intergenic spacer comparison. Comparison between available intergenic sequence elements from XbHW302 and Xlh3 reveal several points of interest. Although H2B and H2A genes show different percentage match/length homologies, the immediate upstream region of each gene does appear to be conserved, suggesting that each of these histone gene types is closely related between the two clusters. Hence the differences on

percentage match/length should not be used as evidence to suggest that genes for the different histone types had diverged at different times, as would be the case if one had evolved by a copying or exchange mechanism involving a completely unrelated cluster type (eg Xlh1 or XbHW61). Rather, it supports the view discussed above, that the rate of silent site change (and so match/length ratios) vary between histone gene types. Busslinger *et al* (1982) noted that early sea urchin H3 and H4 genes showed approximately similar rates of divergence, as was found with the XbHW61 and Xlh1 H3 and H4 genes.

The results of the experiment using SH115 DNA, the *X. borealis* major cluster spacer element, as a southern blot probe (see figure R8) on DNA from various frogs is of interest in this regard. This element appears not to be present in the *X. laevis* clone Xlh3 from the computer search. Has this sequence element been translocated from some other location in *X. laevis*? The southern blot suggests not, and so one must presume that it arose *de novo* in the *X. borealis* lineage, or was lost from the *X. laevis* lineage. If the former were the case, then the formation of SH115 could be analogous to the changes in far upstream sequences in XbHW61, which appear to be recently evolved. The Southern blot in Figure 7b shows that SH115, (or a closely related sequence) is also present elsewhere in the *X. borealis* genome. This could indicate, either the source of this element, or possibly some translocated copy; an orphon-like

element. One should be aware of the limitations of the Southern blot experiments. While this element is high copy number in *X. borealis*, it could be low or single copy number in *X. laevis*, and so miss detection in Figure R8. No single copy control was included, although Figure 7b is a long exposure.

To confirm that all the DNA sequence similarities discussed were specific to the cluster pairs indicated, Xlh1/XbHW302 and XbHW61/Xlh3 comparisons were performed. These revealed no significant homologies. Clearly processes that have not only changed the gene organisation within the clusters, but have also shifted spacer sequences characterise the evolution of these two clusters. The mechanisms involved are considered next.

6.C.3 MECHANISMS OF CLUSTER VARIANT FORMATION

XbHW61/Xlh1 CLUSTERS

The molecular mechanisms which account for the similarities in Xlh1 and XbHW61 are not obvious. Clearly DNA-mediated recombination rather than RNA intermediate gene movement must have occurred, since there is conservation of regions flanking the transcribed parts of these genes. Analysis of the sequence does not reveal multiple repeat elements that could promote unequal crossover events (see below), nor homocopolymers that might produce recombination by the mechanism proposed by Hentchel (1982), nor terminal repeats that characterise the insertion of transposable or

retroviral elements. In fact, it is interesting that the boundaries of conserved and non-conserved sequence are not distinct (figure R19). This could suggest that multiple recombination events have occurred. As the gene order and polarity have not changed, it would appear that these recombination events have been confined to these distal upstream regions. There is no obvious functional advantage from these changes, that could have been selected, which leads one to think that the changes are the result a random nuclear processes associated with cell turnover and repeated sequences.

XbHW302/Xlh3

Clearly in the case of Xlh3 and XbHW302, one has to consider additional mechanisms to account for the greater gene cluster diversity. Some regions upstream of genes display a high degree of similarity between the two clusters, eg the H4 gene. This clearly points to DNA-mediated recombination processes. Although the genetic rearrangements could have evolved between various exchanges between the evolving clusters, the overall differences can be accounted for by a few major excision/inversion/insertion events. Superimposed on this are minor deletions, translocations, and insertions, as indicated on the matrix plots.

For simplicity the overall change from an XbHW302 to a Xlh3 arrangement will be considered. Of particular interest is the change in location of the H2A, H2B gene

pair, together with the spacer sequence downstream of the H1 gene (region 3) of clone XbHW302. This latter element is located in an inverted orientation, adjacent to H2A gene upstream region in Xlh3 relative to XbHW302. This could indicate the excision of a 1.4 Kb fragment. Parts of this fragment appear inverted at a different location, downstream of the H1 gene in Xlh3. This is further supported by the position and polarity of the H2B gene and close upstream region. The inversion of a 1.4 Kb fragment excised from XbHW302 cannot alone account for the formation of the corresponding region of Xlh3, as it is some 5.3 Kb in length. Included in the additional sequence, is the (CA)n repeat, which also appears towards the H3 end of the genic region in XbHW302 (figure R4). However, the sequence at the exact site where the H2A/H2B intergenic region in XbHW302 has split is marked by a HindIII site (see figure A4.4 and A.4.5). This throws doubt on the authenticity of clone Xlh3. This clone was isolated from an partial AluI/HaeIII library (Zernik *et al* 1980). The AluI recognition site encompasses the central four bases within the HindIII recognition site. Thus it is possible that clone Xlh3 is a cloning artifact, ie contains multiple insertions. Certain genomic Southern blot experiments were performed, but these do not tell us if the Xlh3 cluster arrangement does exist in the *X. laevis* genome (Perry *et al* 1985). There are arguments suggesting that multiple cloning might have occurred. First is the occurrence of a site used in the cloning procedure, at the

site of proposed recombination. Secondly, the H2A and H2B arrangement in Xlh3 is unexpected. Excluding sea urchin (1.B.2) and trout (1.B.4), most all other histone gene clusters have the H2A and H2B genes arranged as in XbHW302. Several points argue against multiple insertion. Firstly, a minimum number of three fragments would have to be involved, if one assumes that the H2A and H2B genes were in the XbHW302 arrangement. The probability that two of these three fragments would contain histone gene, is remote, let alone all three. Even more remote would be the chance of cloning one of each histone gene types on three fragments. Furthermore, these fragments would all have to be related to XbHW302, is must be derived from copies of the same (or very similar) cluster types. Another possible explanation, is that Perry *et al* sequenced various subcloned fragments, which they subsequently did not align correctly. Without irrefutable evidence, it is impossible to conclude that Xlh3 is a cloning artifact, but that it may be should be borne in mind.

It is possible from gene polarities that a fragment containing the H4 and H3 genes was inverted. The retention of the H4 flanking sequences, following the change in the gene's polarity supports this hypothesis. It would appear that the evolution of Xlh3 and XbHW302 is best accounted for by several (at least three) major inversions, as well as movement of three types of repeated elements [(AGC)n, (CA)n, and polyT] and the insertion/deletion of

several kilobases of spacer sequence. Presumably, additional to this, minor deletions, insertions, and translocations occurred, as indicated by the matrix plots in figure A4. The complete sequencing of the XbHW302 will further clarify the exact extent of all these movements. In *X. borealis*, it is known that there are 56-63 copies of XbHW302, tandemly arranged. In *X. laevis*, Xlh3 is known to be repeated (Perry *et al* 1985). As considered earlier, repeated elements appear to evolve rapidly. Thus from this analysis, it appears that high copy numbers could promote inversions.

This sequence-based analysis between XbHW302 and Xlh3 shows that cluster topology can be a weak feature with which to assess how closely related two clusters may be. By a few molecular processes, gene orders, intergenic distances and polarities can be completely altered. This renders the attempts to trace histone gene cluster evolution using these features somewhat futile, and explains why such attempts have been fruitless (Hentschel and Birnstiel 1981, Maxson *et al* 1983c). Sequence comparisons between sufficiently closely related species can overcome this problem, because far more individual events (eg base mutations), which can take a longer time to occur, can characterise evolution at the nucleotide level. However, the distal intergenic region of XbHW61 appears anomalous in this respect. It appears to have evolved at the nucleotide level very rapidly, relative to

the divergence of silent sites in neighbouring genes, and proximal 5' regions, possibly due to the existence of several copies in the genome.

Summary

Comparisons between *X. laevis* and *X. borealis* histone gene clusters revealed two pairs displaying differing degrees of structural divergence. The Xlh1/XbHW61 pair can be considered as early intermediates in the gene shuffling process, while the Xlh3/XbHW302 pair are clearly end products; the genes are rearranged. From analysis of the XbHW61/Xlh1 pair, it appears that the first features of cluster evolution can be complete divergence of regions several hundred bases in length, in the far upstream intergenic regions. The mechanisms producing these changes are not clear. Comparison of the XbHW302/Xlh3 pair indicated that several DNA recombination processes, and in particular inversions, had occurred. These mechanisms could be a function of copy number.

6.D REPETITION OF NEW VARIANTS

Section 6.C considered the generation of new cluster variants. In this section, the finding that the major cluster is tandemly repeated (Chapter 4) is considered.

6.D.1 SIZE OF THE MAJOR CLUSTER REPEAT

Unlike all the major cluster clones that were isolated initially, clones XbMW9 and XbMW11 each contain an entire repeat. As these clones were extensively mapped for several restriction enzyme sites, it is possible to determine the size of the repeat by addition of the fragment sizes produced by each enzyme. It is clear that the major *X. borealis* histone cluster has a repeat length of 15.0 Kb.

In the initial characterisation of the major histone cluster of *X. borealis* (Turner & Woodland 1983), there is a suggestion that the minimum length for the cluster is 16.1 Kb, namely the sum of a 12.8 Kb and a 3.3 Kb fragment. There is an explanation as to why this figure contrasts with the results presented here. Firstly, the electrophoresis gels used in the initial investigations were of 0.7 % agarose, which is optimal for the measurement of smaller fragments (1-6 Kb), but could lead to an inaccurate measurement of a 12.8 Kb fragment. The cloned DNA in this study was electrophoresed on 0.7 % agarose, but as the SalI site in the intercluster region was not methylated (see section 4.D), two smaller fragments (4.5 Kb and 7.2 Kb) ran into the higher resolution region of the gel. Secondly, it

is well known that the mobility of DNA is not only dependant on the size of the fragment, but also to a limited extent, dependant on the amount of nucleic acid loaded per track. Typically, 10-100 ng of DNA is radioactively end-labelled, of which only a very small proportion is initially used in a marker lane. The genomic DNA was loaded at 2 ug/track, which would result in these fragments running slightly slower.

By contrast, the mapping of cloned DNA overcomes these problems. Firstly, as mentioned, the fragments tended to have mobilities for which the size could be accurately determined. Secondly, as non-radioactive size markers were used, these were loaded at the same concentration as the restriction digested DNA (about 1 ug/track). Thirdly, size estimation from photographs of cloned DNA tend to be more accurate than measurements of autoradiograph bands produced by Southern blot. These factors could account for the fairly small discrepancy observed.

In this study, genomic DNA was separated on agarose of 0.4 %, which gives an improved and hence more accurate resolution for larger sized fragments.

6.D.2 TANDEM REPEAT COMPARISON

The repeat lengths for those animals with tandemly reiterated histone gene clusters have been tabulated. Table III shows a comparison of repeat lengths.

SPECIES	REPEAT LENGTH	COPY NUMBER	REFERENCE
Sea Urchins		several	Hentschel &
P. miliaris	6.3,6.7	hundred	Birnstiel (1981)
S. purpuratus	6.5		
P. lividus	7.0		
L. pictus	7.2		
Drosophila	4.8,5.0	100	Lifton *et al* (1978)
Artemia	8.5	-	Bagshaw *et al* (1984)
Newt	9.0 +spacer	600-800	Stephenson *et al* (1981)
X. borealis	15.0	55-60	Turner & Woodland (1983) This thesis

Table III. Comparison of repeat lengths and copy numbers from animals known to display tandemly repeated histone genes.

Discussion

It is interesting to note that all the tandem repeats show certain common features. Firstly all the repeats contain single genes for all five histone types. Secondly, most strikingly, tandemly repeated clusters are extremely, but not totally, homogeneous. These features, which are common to tandem repeats in the other species, hold true of *X. borealis*. In contrast to these common features, tandem repeats display certain phylogenetic variations. These include gene order, relative polarity of transcription, distance and sequence of intergenic spacer DNA, and overall repeat length.

Although histone gene organisation has been studied in a wide range of species, only in sea urchin, *Artemia*, and *Drosophila*, has contiguous tandem repetition been shown. The results presented here indicate the fourth animal to display this arrangement. The case of the newt is special, as the clusters, although homogeneous and high in copy number, are separated by long, and possibly variable lengths of satellite spacer DNA (see section 1.B.7). Thus there is no unit length which is contiguous, and reiterated in a tandem fashion.

The comparison in Table III illustrates three points of particular interest. Firstly, *X. borealis* is the most "advanced" eukaryote which has tandemly repeated histone genes. This arrangement was previously commonly associated with the invertebrates. Secondly, *X. borealis* displays the longest tandemly arranged histone gene repeat

noted to date. Thirdly *X. borealis* displays fewer copies of tandemly arranged histone gene repeats, than any other species. The possibility of a link between repeat length and number is considered below.

6.D.2b TANDEM REPETITION IN *X. LAEVIS*

It is worth considering the proposal that *X. laevis* histone genes are tandemly repeated. Originally the arrangement of sea urchin histone genes was the paradigm of histone gene organisation. This arose because many fewer of the heterogeneous higher eukaryote histone genes had been investigated. Zernik *et al* (1980) isolated the clone Xlh1 from a *X. laevis* genomic library, using an H4 cDNA gene hybridisation probe. This clone contained seven histone genes on a 15 Kb fragment. The initial characterisation of the histone gene order was based on hybrid release translation. This yielded the gene order H4-H3-H1-H2B-H2A-H4-H3, which appears to suggest that the clone contained five histone genes from one repeat, and two from the neighbouring tandemly repeated cluster. However analysis of the restriction sites in the regions at either end of the clone displayed gross heterogenaity, which showed that the case for tandem repetition, as understood for sea urchin genes, was not valid (Hentschel & Birnstiel 1981). More recently the same group has yielded more data which confirms the objection noted by Hentschel and Birnstiel.

Sequence data of the regions between the H3 and H4 at either end of X1h1-H1, shows sequence divergence (Perry *et al* 1985).

Destree *et al* (1984) described the gene localisation and restriction site maps of a set of three clones (X1-hi-102, X1-hi-103, X1-hi-105) which, when aligned, appear to display two neighbouring, tandemly repeated clusters. A fourth clone, (X1-hi-118) displayed a different gene arrangement. This clone contained two genes of each histone class, except that it lacked a second H1 gene. The genes were arranged into two neighbouring clusters, which could be interpreted as representing part of a tandemly arranged block of clusters. However, both these different cluster types displayed restriction site patterns which dismiss the possibility of regular tandemly repeating unit. In this respect they are similar the clone X1-hi-1. This would tend to suggest that instead of any of *X. laevis* cluster types mentioned representing part of a tandemly arranged 'array' of clusters, each repeat may occur at different loci. This would account for the heterogeneous restriction site positions outside the genic regions of each cluster type.

More recently, Perry's group has reproposed tandem repetition, but for the X1h1 cluster type, and with very scant evidence (Perry *et al* 1985). In this case, the suggestion is that the majority of the *X. laevis* histone genes are present in a single major cluster. An substantial

arguments against this specific point have been made (Turner & Woodland 1983), it is surprising that it has been reproposed. This group contend that the clone Xlh1 is not the most predominant cluster type, although other evidence suggests it may be (Old *et al* 1982). Section 1.B.8 explains that Southern blots on DNA pooled from many individuals will display the structure most common to many organisms, though not necessarily the most common in any single organism.

Southern blotting of pooled genomic DNA digested with ScaI revealed a 15 Kb fragment when either an H1 or H4 histone gene hybridisation probe from Xlh1 was used. Perry *et al* suggest that Xlh3, rather than Xlh1 is the predominant, or major cluster. As digestion with other restriction enzymes yields few predominant bands, Perry *et al* conclude that this cluster is tandemly repeated. The evidence falls short of being convincing because, unlike the case of *Artemia*, there is no evidence that the other predominant restriction fragments span the ScaI sites, a second condition in order to confirm tandem repetition. Thus it is perfectly possible for the present evidence to indicate a common or predominant cluster type or repeat, which is bordered by ScaI sites, but yet dispersed throughout the genome. No mention of this possibility is given by the authors. Such conserved, but yet dispersed cluster arrangement is not unprecedented. The rainbow trout, which, though it is not closely related to *Xenopus* has a similar developmental strategy (M.R.Woodland pers. comm.),

displays exactly this arrangement (see section 1.B.6), and there is other suggestive evidence for this arrangement in *X. laevis*. Destree *et al* (1984) isolated four *X. laevis* clones (Xl-hi-104, Xl-hi-106, Xl-hi-110, Xl-hi-114) which contain this cluster, or part of it. None of these clones show consistent restriction site positions in the spacer at one end of the cluster, as would be expected for true tandem repetition. Unfortunately Destree *et al* did not map the ScaI sites on their isolates, nor, due to the differing choice of enzymes used to map this cluster, do the maps produced by Destree *et al* disagree with those of Perry *et al*.

The work presented in this thesis shows the existance of a tandem repeat structure in *X. borealis* which is similar to the *Drosophila* and sea urchin type in that there is very little restriction site heterogeneity. As consideration of *X. laevis* derived clones show restriction site heterogeneity, it would seem that the major *X. laevis* clusters are not tandemly repeated in the fashion that the major cluster of *X. borealis* is. A chromosome walk would resolve this question.

6.D.3 CONCERTED EVOLUTION

Since the first discovery of the tandemly repeated and highly homogeneous sea urchin and *Drosophila* histone genes, there has been interest as to how such clusters evolve. This evolution results in different species displaying structurally different tandemly repeated histone gene

clusters, each highly homogeneous. Analysis has shown this high intra-specific homogeneity of tandemly repeated genes is not confined to histone genes. It has also been observed in the 28S and 18S ribosomal DNA, 5S rDNA, heat shock genes, and repeated sequences of unknown function (For reviews see Jeffreys 1982, Long & Dawid 1980, Fedoroff 1979, Dover et al 1982)

As the homogeneity appears to be maintained in the species, it has been inferred that a continual process of homogenisation occurs. The effect of this has been termed 'Concerted Evolution', as each member appears to evolve in concert. The term is perhaps not ideal, as it suggests parallel evolution of each member independently, rather than simply an observation of family homogeneity. Concerted evolution could therefore be termed as the processes which establish and maintain greater intra-specific homogeneity of repeated genes than inter-specific homogeneity. These multigene families or repeated sequences appear to evolve uniformly, and so one is led to presume there must be some feature of these sequence elements which allows or causes this process. It is clear in the case of the *X. borealis* major histone gene cluster, a new variant was formed. This variant became fixed and was multiplied up to form a battery of tandem repeats. Clearly this new arrangement must have also spread rapidly through the population, as it is found in every individual studied to date. Before considering the

possible mechanisms which could bring about concerted evolution, it is worth considering some other features, and consequences of this process.

The occurrence of concerted evolution has prompted research aimed at characterising the extent and rate of this process. There is evidence that the *X. borealis* major cluster lies on one particular chromosome (Turner *et al* 1988). However homogenisation between arrays is believed also to be able to occur between different chromosomes. Comparisons of ribosomal gene DNA (18S and 28S genes) in seven sibling species of *Drosophila* show that fixation of a variant had occurred in parallel on both X and Y chromosomes (Coen *et al* 1982). A second point of interest is the rate of concerted evolution. Assuming that the mechanism of concerted evolution involves the formation of a single variant which becomes fixed by whatever means, one may divide the rate of evolution of each region into the rate of change and the rate of fixation. To study this process, the sequence similarities of ribosomal genes in seven sibling *Drosophila* species were investigated (Coen *et al* 1982). Clearly, repeated sequence elements with a low rate of change are more highly conserved, and those of a high rate, more variable. Non-transcribed ribosomal DNA (NTS), which has tandemly repeated elements in it, displayed the most variation. The transcribed spacer between the genes and the transcribed spacer outside the genes were respectively increasingly more highly conserved. The rDNA coding region

appeared the most conserved and was homogeneous between all species. As between all seven sibling species the NTS regions displayed different fixed variants, fixation must have occurred between the time of divergence of the most similar species. This shows the rate of fixation for this element at least, to be evolutionarily rapid. One would predict a significantly greater rate of fixation than rate of change, or the high level of homogeneity could never arise.

The discovery of the *X. borealis* major cluster, and its absence in *X. laevis* has similarities to the above study. Clearly the time for fixation of the major cluster variant can not exceed the time since these species diverged. This also holds true for the minor cluster diverged region. Clearly the non-transcribed spacer sequence in *Drosophila* ribosomal genes (as the 5S genes in *Xenopus*) has diverged completely, ie more than even the XbMW302/Xlh3 pair. The hypothesis that the rate of spacer evolution is a function of copy number is supported by this.

The formation of the major cluster in *X. borealis* raises question as to why no similar event has happened in *X. laevis*. Possibly this process could be currently occurring in *X. laevis*. Possibly the *Xenopus* populations are larger and more complex, and that different sampling would yield different results. Could it be that the forces

creating cluster diversity in *X laevis* have acted faster than fixation forces in *X. laevis*, while the opposite is the case in *X. borealis*.

It is clearly possible that while mechanisms that generate cluster diversity may act frequently (in evolutionary terms), as indicated by the diversity of histone gene clusters in *X. laevis*, the fixation rate of one variant cluster may be triggered by a single event. Such single events could occur with a low probability, such that, by chance, it happened once in the *X. borealis* lineage, but not in the *X. laevis* lineage. Conversely, there could be feature(s) specific to the *X. borealis* major cluster, that promote such events. From detailed comparative analysis between Xlh3 and XbMW302, it appears that at least five DNA exchange events occurred to produce the observed differences. It does appear that Xlh3 is repeated, within the *X. laevis* genome (Perry *et al* 1985). Possibly *X. laevis* and *X. borealis* multiplied up different elements that existed in a common ancestor (Turner & Woodland 1983). Naturally the rate of change of different element types can be explained by differences in plasticity of structure which are allowable without detrimental consequences on biological function.

The fixation rate of a variant is thought to be comprised of the rate of homogenisation of a variant within an array on a chromosome, the rate of fixation of that chromosomal array in a population, and the rate of transfer

of the variant units between chromosomes (Smith 1976, Coen *et al* 1982). The fixation rate is believed to be dependant partly on the number of repeats.

Clearly generation time, and population sizes will affect the time taken for a neutral variant to spread though the population. Although some of these components may be known for *X. borealis* (ie copy number and generation time), details of population sizes are much harder to estimate, and so error factors in any calculations may be so massive as to render such calculations meaningless. As *Xenopus* has a generation time which is much greater than *Drosophila*, the rate of spread throughout the population may be expected to be at least proportionally longer, all other things being equal.

6.D.4 MECHANISMS OF CONCERTED EVOLUTION

There are several mechanisms which may explain how the effects of concerted evolution occur. By far the most favoured explanation for the generation of tandem repeated homogeneous histone gene clusters is that of unequal crossover (Coen *et al* 1982, Dover *et al* 1982, Hentschel & Birnstiel 1981, Kedes 1979).

Three consequences of this are pertinent to the generation of tandem repeats in *X. borealis*. Firstly, the number of repeats. The first unequal cross-over will produce a single pair of tandem repeats. If there were originally two regions of homology, the first cross-over will result in three regions of homology. If a second unequal cross-over event occurs, there will be the possibility, if the most distant sites are used, to double the tandem pair. Naturally, subsequent events can, at maximal rate, double the number of repeats each time, and hence generate large copy numbers. Secondly as this process can in theory double the numbers of repeats in one event, it may help account for the observed rapidity of the formation of the *X. borealis* major cluster. Thirdly, the process produces repeats of high homogeneity. In *X. borealis* there are a predicted 56-63 copies of the major cluster. In this thesis, evidence has been presented showing that the majority of the major cluster repeats are tandemly repeated. In theory to generate enough tandemly arranged copies at the maximal rate, six (2^6=64) unequal cross-over events are required. As the

minimum possible number, this may nevertheless seem a very large number of what may be thought of as a rare chance event, but as illustrated with the consideration of the second event above, more homology sites are generated. As the whole process is dependant on regions of homology, the chance of such an event occurring should increase with each event. Clearly there must be forces, possibly selective, which act to limit the spread of this process throughout the genome, or at least along chromosomes.

This process can also be used to explain the origin and behaviour of repetitive DNA in general. As mentioned, this process can occur between homologous chromatids at mitosis, or homologous chromosomes (non-sister chromatids) at meiosis. There are many more mitoses than meioses in the germ line, and the number of chiasmata observed at meiosis roughly equals the number of sister chromatid exchanges observed at mitosis (Smith 1976). So it seems reasonable to suppose that the number of mitotic sister chromatid exchanges per generation occurring in the lineage leading to a given chromosome, greatly exceeds meiotic non-sister chromatid exchanges. There is still no direct evidence for such recombination, however strong support comes from studies of yeast ribosomal genes. Use of genetic techniques unique to fungi, allowed the measurement of unequal cross-over frequency in yeast. It was shown that unequal crossover could account for the concerted evolution of this family (Petes 1980, Szostak 1980).

Discussion

These results can be extended to account for the sequence homogeneity of tandem repeated genes, such as the histone gene family. This sheds light on a possible reason for clustering. The histone genes are not transcribed in a polycistronic fashion (Portman *et al* 1976). This raises the question of why they should be clustered. In yeast where H4 and H3 genes are on different chromosomes to the H2A and H2B genes, clustering is not required for coordinate expression. Although little is known about how coordinate expression is regulated, the maintenance of approximately equal numbers of genes could be important. In a situation where all the histone genes are situated close together (ie clustered) the relative number of each histone gene type remains balanced after an unequal cross-over, provided that the regions of homology exist outside the cluster structure. Thus if multiplication is favoured, it would also be multiplication of a quintet monomer that would be favoured. Unequal cross-over therefore appears an attractive mechanism to account for the rapid formation of the homogenous tandemly repeated major cluster in *X. borealis*.

It may be considered that the large (15Kb) repeat size in *X. borealis* is likely to reduce the chance of mis-alignment interactions of homologous regions, relative to say the sea urchin with a repeat size of 6-8 Kb. This may account for the lower reiteration frequency (55-60) of *X. borealis* relative to the sea urchin (several hundred), although the differences in developmental strategy probably

play a greater role (see Woodland *et al* 1983).
Alternatively, the major histone repeat in *X. borealis* may have only recently been fixed, and may increase in copy number before stabilisation.
While on the topic of unequal cross-over, it is interesting to consider *N. viridescens*, which displays many more (600-800) copies of apparently identical histone gene clusters. One would predict that to sustain the apparent homogeneity of sequence, unequal cross-over events would have to occur frequently relative to the rate of change. On the basis that the frequency of regions of homology may determine the frequency of unequal cross-over events, the histone genes of *N. viridescens* are well placed for very frequent unequal cross-over events. For although these histone gene clusters are 9 Kb in length, they are embedded in highly repetitive satellite DNA.

Other possible mechanisms to account for the generation of histone gene clusters, and the rapid flux in organisation between so closely related species, are not so appealing.

Gene conversion, a process which has been observed both within a chromosome (Klein & Petes 1981) and on separate chromosomes (Scherer & Davis 1980), does not require tandemly arranged DNA.
Thus it can account for recombination changes affecting members of dispersed, irregularly arranged histone gene families. From comparisons between the two copies of each

core histone gene in yeast, gene conversion rates have been deduced to be very low (Smith 1984). However as the copy number in *Xenopus* is so much greater, it is unwise to conclude that gene conversion is too infrequent to play a role in generating diversity.

The interest in transposable elements as possible vectors in histone gene cluster evolution has been strengthened by discovery in *Drosophila* of elements of the family copia-297 which appear to specifically intergrate in the TATA boxes of H3 genes (see Maxson *et al* 1983c). A sea urchin transposable element, the TU1 element which inserts in the H2B gene at position 57 (Weinthal *et al* 1983), has the characteristic terminal inverted repeats. Presumably any movement of DNA, by transposable elements may effect molecular evolution.

Horizontal gene transfer is a process about which little is known, but is quoted as possible to account for hitherto inexplicable or highly unlikely occurrences. The case presented by Busslinger *et al* (1980) to account for very similar sequences between histone clusters in two species of sea urchin falls into this category. There must be an explanation of the homology in this case, and whatever it is, it too may play a role in changing certain aspects of histone cluster structure, while retaining others, and could occur in *Xenopus*. Interestingly, the *X borealis* major cluster displays the same gene order and polarity as *Drosophila*.

Lastly, a further mechanism by which recombination events can occur has been suggested by Hentschel (1982). Analysis of S1 nuclease sensitive sites of sea urchin histone genes cluster repeats revealed homocopolymer sequences. These sites which are within each repeating unit appear to flank the H1 gene. It has been proposed that out-of-register DNA slippage could occur in these regions, which would leave single stranded structures. These could then be foci for recombination events, unequal cross-over events, or gene conversion events. From analysis of the *X. borealis* major cluster interganic sequences (Figure A1), it is apparent that both at either end of the *X. borealis* ganic region, and between genes, homocopolymers exist. These could possibly explain the means by which the histone gene clusters have been able to evolve rapidly in *Xenopus*.

6.D.5 MOLECULAR DRIVE

The observation that multigene families and repeated sequences display concerted evolution has lead to the term 'Molecular Drive' (Dover 1982, 1986). Molecular Drive has been defined as the fixation of variants in a population as a consequence of stochastic and directional processes of family turnover. The stochastic component refers to random fixation of a variant, and is likely to be caused by unequal crossover, gene conversion and transposition. The mechanisms for the directional component that have been proposed are duplicative transposition and biased gene conversion. They

appear to promote the frequency of specific sequence elements. Biased gene conversion favours the change to one particular gene. Duplicative transposition is thought to occur, not by the physical excision and re-insertion of a DNA sequence, but by copying of a donor sequence, which remains unchanged, while the copy sequence inserts elsewhere. Loss by excision would occur as an independant process. (Spradling & Rubin 1981)

From the discussion in section 6.C, it appears that DNA inversions and insertions could account for the differences in histone cluster structure between *X. laevis* and *X. borealis*. Unequal crossover could account for the tandem repetition of the *X. borealis* major cluster. These alone can not account for molecular drive. The fact that the major cluster and the far upstream element in the minor cluster, appears to be reproduced in every copy within the individual, and across the species, indicate that mechanisms working in addition to those proposed, rapidly spread the changes through the population.

Molecular Drive has been suggested to have a broader role in speciation. It appears that the fixation of variants in a repeated sequence family can spread throughout the population at a rate greater than the natural divergence of these sequences. Hence, members of a panmictic population would tend to evolve together. Increasing the similarity of

such a population will tend to decrease similarities between populations, which could, if the changes adversely affected the hybrid fitness, lead to speciation.

The factors which caused the speciation of *X. laevis* and *X. borealis* are unknown, but differences in histone gene cluster structure could have arisen in different populations of a common ancestor. These changes could have had an adverse affect on the fitness of individuals resulting from mixed matings between these populations. Selection against such crosses may in turn have promoted the formation of the two species. Could the formation of the *X. borealis* major cluster have been that change? *In situ* hybridisation studies showed that an H4 gene probe binds to several loci within the *X. laevis* genome. The same probe used on *X. borealis* chromosomes revealed the predominant hybridisation to the terminal section of one particular chromosome (Turner *et al* 1988). This presumably is the major cluster, as the major cluster is known to carry 70% of the H4 genes, and as is shown in this thesis, the major cluster is tandemly repeated. During meiosis, homologous chromosomes have to align. The estimated 56-63 copies of a 15 Kb repeat (840-960 Kb DNA) may have been unable to align with the homologous chromosome in mixed matings between different ancestral frog populations, where one population had evolved the major cluster, and the other had not. This would then inhibit meiosis, and cause infertility. Such infertility would promote speciation.

Discussion

Naturally in practise there are far more numberous and likely causes of the speciation of *X. laevis* and *X. boreallis*, however if this was the cause of speciation, one might expect that crosses between *X. laevis* and *X. boreallis* would be viable, but sterile, which they are.

Upstream H1 gene region

gggatcaaca gggcctattc ctctactaac tttccagaac cttcctgtat cgaagcacat tcgaccaggg ggcaagcagc
ttcttgggcg gccgaagcag caggggttcc tgtggatgtg atagtcccgg cagcatcctg gtctacgcca tgcacgttca
tttcccactg ccaactagac ctcaggagtt caggagctct cctctttggt tccagaatcc tgtcctctgc ttcttcttct
tctcctacac attaaaatgt atggcatctt atcagtctca gatatattat tcccacccgg ggatgactac cttgctagct
ccgggagact attgacatcg cgaggactag ggaaaggga gaattgtttc at Approx 1Kb to H1 gene

H1 gene region

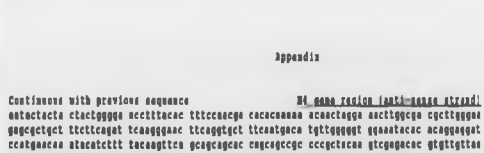
agctcccttc atgctcgatg ggcaaagtag atgacgggcg tggctacgtt ttatccaatc accaaagttt tgcaaatgaa
acggaccaa tcaacaagac gcaacacgtt cttatataaa gagaagcgga aagccgccag atcatatttt tctttgttaa
aacgacgagg natttgctaa tggctgaagc cacccatgcc gcccccgctc ctcccccggc tcaacccgca gccaagacaa
agaaacagcc gaagaaaccg gtgcccccgc ctagctccaa gaagcccgct gggccaagtc tctcccaact gatccgtcaa
gcccgtatcc ctccgcaaga gcgcagccgg gtgtcccctg caccccctca gaagcctcta gaagcctctc gctgctggag gatacgatgt
ggacaaaaat aacagtccac tcaagctgac tctcaagcac ttggtcagca acgaagccct tgtccaagtc aaagccgcgc
gaccctccgc cyccttcaag ctcaacaaga aacagctgca gaccaagcac aagtcgccca acgacaagcc accgctgccc
aacaagccag cggccaaaaa gccagccgca aagtccccga aaaagcccaa gaagctctac gcgccccgca agaccccaaa
gaagctccag aaaccccccg aggccgccag caagaaccca aaaaaacccg aagccgccaa gcccgaagaa gtgcccgcag
gcccccgcaa aaagaagcgt caagccaaag ctgccaaaag ccccgccaag gccaagcag ccaaaccca agtgccgcaa
gcaagcaaag ccgccccgaa gaacaaagtga gcaccgtcac tactagtggc caactgaacc aaaaactctt ttaaaaacca
ccaactccct ccccaacgtg cttctcctct taactgccc gagtctccty cttctaacac acgttttcct actcaactct
tttctccctt gttgttgttg tagtagtagt agcagcagca gcagcaatag tagtagagtt aaccaacatg tagtcgtttg
gcggtgccca ttgccgccca cagtgaggga gaaccttctc ttcgggaggc ggagacgcta catggaccga catggcctcc
cttgacgggg atcttttcag aagaaatatt tagtccggta aaccaactgg cagagtagca cagagtgcca cagagctcag cggagaggct
catggcttg tgctactagt gcggtcctgg gggaaacggg aaagacgtag aagatcaagg tagcgatttc gtgccaaggc
tggccaaaag aaataaatgg cggaaaccct gcagaaagtc tatcc

H3A gene (anti-sense) and H2A gene regions

gcaaggaagg taggATTACT TCGACTCGCT GTATTTGGTC ACACCCTTGC TCCCCTCCGA CACGGCGTGC TTGCCCACCT
CCCCCAGCCAG CAGCAGGCGG ACAGCAGTGT CGATCTCCCG GCAGCTGATG GTGCAACCCT TGTTGTAGTC AGCCAGCCGG
CAGCCTCCCC CGGCGATGCG CTCGACACCA TCGTTACCAA AGCAGTTCAT GATCTCATC GCCTTGGAAG AGATCCCGCT
ATCAGGGTGC ACCTGCTTCA GCACTTTGTA CACGTAGATG GCATAGCTCT CCTTCCTGCT CTTCCTACGC TTCTTGCCAC
TTTTCTTCTG GGTTTAGTC ACGGCTTTCT TAGAGCCCTT CTTTGGGGCC GGAGCGAACT TGGCAGGTTC AGGCATattc
agagactaga caagatcaaa cagcgagcga gcgctaatan agcaatgctg gtccgtccct gcagcggctt ttatagcctg
cccccatgcn aagcagcttg cgctggtacn ctccttcccn ttggctggct ctgtagcatg acgtccctti cttgcgtttg
cgtgaagacg atcggcggta agtaagcgtt gctgttcgtt ggctgcgagg aagccttaag CAATcggcgn ggggcgagtn
gcgtggcagc tacctataaa aggcaaagtg cgcgctccaa agtattttct gaagtcactt tgtgtccaca gcgaaggctt
ttcatccatg TCAGCAAGAG GCAAACAAGG AGGCAAGCCC CGGGCTAAGG CCAAGACTCG CTCATCTCGG GCAGGCTTGC
AGTTCCCAGT GGGCCGTGTT CACCGGCTGC TCAGCAAGGG CAATTACCCT GACCGGGTGG GAGCTGGAGC CCCATCTCAC
CTTACCGCGG TACTCCAGTA TCTCACCGCT GAGATCCTGG AGTTGGCCGG CAACGCTGCC CGGGATAACA AGAAAACCCG
TATCATTCCC AGGCACCTGC AGCTCGCCGT GCGCAATGAC GAGCAGCTCA ACAAACTGCT CGGAGGAGTC ACCATCGCTC
AGGGCAGTGT CCTGCCCAAC ATCCAGTCCG TCTTGCTCCC CAAGAAAACC GAGAGCGTCA AATCGGCCAA AAACAATCGA
gcctccagcg acacaacgca acctctgcc ctcaccaccc acccaaccnn ctctttcaa aaccaccact gtctctccaa
gcgctgttct ccgcttgct tctatcactg cgcagtgtcn ttgccgcgcc tcccccttct ctttttcgg ccattacttg
acgccaccgt gaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa

Figure A1(a) : The X.borealis major cluster clone XbMN102 sequence. See figure A2
The coding sequences are indicated in capitals.

Continuous with previous sequence — B4 gene region (anti-sense strand)



B3 gene region

Downstream B3 gene region

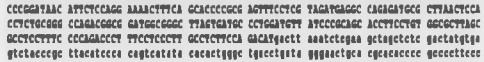


Figure A1(b) : The *X.borealis* major cluster clone XbW302 sequence. See figure B2 The coding sequences are indicated in capitals.

```
                                                                  10                                                                  60
ATG GCT GAA GCC ACC GAG TCC GCG CCC GCT CCT CCC CCG GCT GAA CCC GCA GGC AAG AAG
Met Ala Glu Ala Thr Glu Ser Ala Pro Ala Pro Pro Pro Ala Glu Pro Ala Gly Lys Lys

                                                                  70                                                                 120
AAG AAA CAG CCG AAG AAA GCG GTG GGG GCC GCT AAG TCC AAG AAG CCC GCT GGG CCA AGT
Lys Lys Gln Pro Lys Lys Ala Val Gly Ala Ala Lys Ser Lys Lys Pro Ala Gly Pro Ser

                                                                 130                                                                 180
GTC TCC GAA CTG ATC GTC AAA GCC GTG TCC GCT TCC AAG GAG CBC AGC GGG GTG TCC CTG
Val Ser Glu Leu Ile Val Lys Ala Val Ser Ala Ser Lys Glu Arg Ser Gly Val Ser Leu

                                                                 190                                                                 240
GCA GCC CTC AAG AAG GCT CTG GCT GCT GGA GGA TAC GAT GTG GAC AAG AAT AAC AGT CGA
Ala Ala Leu Lys Lys Ala Leu Ala Ala Gly Gly Tyr Asp Val Asp Lys Asn Asn Ser Arg

                                                                 250                                                                 300
CTC AAG CTG GCT CTC AAG GGC TTG GTC AGC AAG GAG ACC CTT GTC CAA GTC AAA GGG AGC
Leu Lys Leu Ala Leu Lys Gly Leu Val Ser Lys Glu Thr Leu Val Gln Val Lys Gly Ser

                                                                 310                                                                 360
GGA GCC TCC GGC TCC TTC AAG CTC AAC AAG AAG CAG CTG CAG ACC AAG GAG AAG GTC GCC
Gly Ala Ser Gly Ser Phe Lys Leu Asn Lys Lys Gln Leu Gln Thr Lys Glu Lys Val Ala

                                                                 370                                                                 420
AAG AAG AAG GCA CCG GTA GCC AAG AAG CCA GCG GCC AAA AAG CCA GCG GCA AAG TCT CCG
Lys Lys Lys Ala Pro Val Ala Lys Lys Pro Ala Ala Lys Lys Pro Ala Ala Lys Ser Pro

                                                                 430                                                                 480
AAA AAG CCC AAG AAG GTC TCC GCG GCC GCC AAG AGC CCA AAG AAG CTC AAG AAA CCC GCC
Lys Lys Pro Lys Lys Val Ser Ala Ala Ala Lys Ser Pro Lys Lys Leu Lys Lys Pro Ala

                                                                 490                                                                 540
AAG GCG CCA GCC AAG AGC CCG AAA AAG CCC AAA GCC GCC AAG CCC AAG AAG GTG GCC AAG
Lys Ala Pro Ala Lys Ser Pro Lys Lys Pro Lys Ala Ala Lys Pro Lys Lys Val Ala Lys

                                                                 550                                                                 600
AGC CCC GCA AAA AAG AGC GTC AAG CCC AAA GCT GCC AAA AGC CCC GCC AAG GCC AAA GCA
Ser Pro Ala Lys Lys Ser Val Lys Pro Lys Ala Ala Lys Ser Pro Ala Lys Ala Lys Ala

                                                                 630
GCC AAA CCC AAA GTG GCC AAA GCA AAG AAA GCC GCC CCG AAG AAG AAA TGA
Ala Lys Pro Lys Val Ala Lys Ala Lys Lys Ala Ala Pro Lys Lys Lys End
```

Figure 22 (a) : the translation product sequence of the H. borealis major cluster histone H1 gene

```
                                                30                                                  60
ATG CCT GAA CCT GCC AAG TCC GCT CCG GCC CCA AAG AAG GGC TCT AAG AAA GCC GTG ACT
Met Pro Glu Pro Ala Lys Ser Ala Pro Ala Pro Lys Lys Gly Ser Lys Lys Ala Val Thr

                                                90                                                 120
AAA ACC CAG AAG AAA AGT GGC AAG AAG CGT AGG AAG AGC AGG AAG GAG AGC TAC GCC ATC
Lys Thr Gln Lys Lys Ser Gly Lys Lys Arg Arg Lys Ser Arg Lys Glu Ser Tyr Ala Ile

                                               150                                                 180
TAC GTG TAC AAA GTG CTG AAG CAG GTG CAC CCT GAT ACC GGC ATC TCT TCC AAG GCC ATG
Tyr Val Tyr Lys Val Leu Lys Gln Val His Pro Asp Thr Gly Ile Ser Ser Lys Ala Met

                                               210                                                 240
AGC ATC ATG AAC TCC TTT GTT AAC GAT GTG TTC GAG CGC ATC GCC GGG GAA GCT TCC CGC
Ser Ile Met Asn Ser Phe Val Asn Asp Val Phe Glu Arg Ile Ala Gly Glu Ala Ser Arg

                                               270                                                 300
TTG GCT CAC TAC AAC AAG CGC TCC ACC ATC ACC TCC CGG GAG ATC CAC ACT GCT GTC CGC
Leu Ala His Tyr Asn Lys Arg Ser Thr Ile Thr Ser Arg Glu Ile Asp Thr Ala Val Arg

                                               330                                                 360
CTG CTG CTG CCT GGG GAG CTG GCC AAG CAC GCC GTG TCC GAG GGC ACC AAG GCT GTG ACC
Leu Leu Leu Pro Gly Glu Leu Ala Lys His Ala Val Ser Glu Gly Thr Lys Ala Val Thr


AAA TAC ACC AGC TCC AAG TAA
Lys Tyr Thr Ser Ser Lys End
```

Figure A2 (b) : The translation product sequence of the H. borealis Major cluster histone H2b gene

```
                                                30                                                  60
ATG TCA GGA AGA GGC AAA CAA GGA GGG AAG ACC CGG GCT AAG GCC AAG ACT CGC TCA TCT
Met Ser Gly Arg Gly Lys Gln Gly Gly Lys Thr Arg Ala Lys Ala Lys Thr Arg Ser Ser

                                                90                                                 120
CGC GCA GGG TTG CAG TTC CCA GTG GGC CGT GTT CAC CGC CTG CTG AGC AAG GGC AAT TAC
Arg Ala Gly Leu Gln Phe Pro Val Gly Arg Val His Arg Leu Leu Arg Lys Gly Asn Tyr

                                               150                                                 180
GCT GAG CGC GTG GGA GCT GGA GCC CCA GTC TAC CTT ACC GCG GTA CTC GAG TAT CTC ACC
Ala Glu Arg Val Gly Ala Gly Ala Pro Val Tyr Leu Thr Ala Val Leu Glu Tyr Leu Thr

                                               210                                                 240
GCT GAG ATC CTG GAG TTG GCC GGC AAC GCT GCC CGC GAT AAC AAG AAA ACC CGT ATC ATT
Ala Glu Ile Leu Glu Leu Ala Gly Asn Ala Ala Arg Asp Asn Lys Lys Thr Arg Ile Ile

                                               270                                                 300
CCC AGG CAC CTG CAG CTC GCC GTG CGC AAT GAC GAG GAG CTC AAC AAA CTG CTC GGA GGA
Pro Arg His Leu Gln Leu Ala Val Arg Asn Asp Glu Glu Leu Asn Lys Leu Leu Gly Gly

                                               330                                                 360
GTC ACC ATC GCT CAG GGC GGT GTC CTG CCC AAC ATC CAG TCC GTC TTG CTC CCC AAG AAA
Val Thr Ile Ala Gln Gly Gly Val Leu Pro Asn Ile Gln Ser Val Leu Leu Pro Lys Lys

                                               390
ACC GAG AGC GTC AAA TCG GCC AAG AGC AAG TGA
Thr Glu Ser Val Lys Ser Ala Lys Ser Lys End
```

Figure A2 (c) : The translation product sequence of the H. borealis Major cluster histone H2a gene

```
                                    10                                  60
ATG TCT GGA AGA GGC AAG GGA GGA AAG GCT CTG GGG AAA GGA GGC GCT AAG CGC CAC AGC
Met Ser Gly Arg Gly Lys Gly Gly Lys Gly Leu Gly Lys Gly Gly Ala Lys Arg His Arg

                                    70                                  120
AAG GTG CTG CGC GAT AAC ATC CAG GGC ATC ACT AAG CCC GCC ATC CGC CGT CTG GCC CGC
Lys Val Leu Arg Asp Asn Ile Gln Gly Ile Thr Lys Pro Ala Ile Arg Arg Leu Ala Arg

                                    130                                 180
AGA GGT GGA GTT AAG CGC ATC TCT GGC CTC ATC TAC GAG GAA ACT CGC GGG GTG CTG AAA
Arg Gly Gly Val Lys Arg Ile Ser Gly Leu Ile Tyr Glu Glu Thr Arg Gly Val Leu Lys

                                    190                                 240
GTT TTC CTG GAG AAT GTT ATC CGG GAC GCC GTC ACC TAC ACC GAG CAC GCC AAG AGG AAG
Val Phe Leu Glu Asn Val Ile Arg Asp Ala Val Thr Tyr Thr Glu His Ala Lys Arg Lys

                                    250                                 300
ACC GTG ACC GCT ATG GAT GTG GTC TAT GCT CTC AAA CGT CAG GGC CGC ACT CTC TAC GGT
Thr Val Thr Ala Met Asp Val Val Tyr Ala Leu Lys Arg Gln Gly Arg Thr Leu Tyr Gly


TTC GGA GGT TAA
Phe Gly Gly End
```

Figure A2 (d): The translation product sequence of the L. borealis Major cluster histone H4 gene

```
                                    10                                  60
ATG GCT CGT ACC AAG CAG ACC GCC CGT AAA TCC ACC GGC GGC AAG GCT CCC CGC AAG CAG
Met Ala Arg Thr Lys Gln Thr Ala Arg Lys Ser Thr Gly Gly Lys Ala Pro Arg Lys Gln

                                    60                                  120
TTG GCC ACC AAG GCA GCT CGC AAG AAC GCC CCC GCC ACT GGC GGA GTC AAG AAA CCC CAT
Leu Ala Thr Lys Ala Ala Arg Lys Asn Ala Pro Ala Thr Gly Gly Val Lys Lys Pro His

                                    150                                 180
CGT TAT CGC CCG GGC ACC GTG GCT CTC CGT GAA ATC CGC CGC TAC CAG AAG TCC ACC GAA
Arg Tyr Arg Pro Gly Thr Val Ala Leu Arg Glu Ile Arg Arg Tyr Gln Lys Ser Thr Glu

                                    210                                 240
TTG CTC ATC CGA AAA CTG CCC TTC CAG CGT CTC GTC CGG GAG ATC GCA CAG GAT TTT AAG
Leu Leu Ile Arg Lys Leu Pro Phe Gln Arg Leu Val Arg Glu Ile Ala Gln Asp Phe Lys

                                    270                                 300
ACC GAT CTC AGG TTC CAG AGC TCA GCC GTT ATG GCT CTG CAG GAG GCC AGC GAA GCT TAT
Thr Asp Leu Arg Phe Gln Ser Ser Ala Val Met Ala Leu Gln Glu Ala Ser Glu Ala Tyr

                                    310                                 360
CTG GTC GGA CTC TTC GAG GAC ACC AAC CTG TGC GCT ATC CAC GCC AAG AGA GTC ACC ATC
Leu Val Gly Leu Phe Glu Asp Thr Asn Leu Cys Ala Ile His Ala Lys Arg Val Thr Ile

                                    390
ATG CCC AAG GAT ATC CAG CTG GCC CGC AGG ATC CGA GGG GAG AGG GCT TAA
Met Pro Lys Asp Ile Gln Leu Ala Arg Arg Ile Arg Gly Glu Arg Ala End
```

Figure A2 (e): The translation product sequence of the L. borealis Major cluster histone H3 gene.

```
Xlh1 h1prot
Xlh3 h1prot

1       MetAlaGluAlaAlaProGlyGlnAlaProAlaProProProAlaAlaGluProAlaGlyLysLys
          |  |  |  |  :  |  |  |  |  |  |  |  |  |  |  |     |  |
1       MetAlaGluAlaAlaGluGlnAlaProAlaProProProAlaAlaGluProAlaAlaLysLys

21      LysLys   GluProLysLys         AlaValGlyAlaAlaLysArgLysLysPro
          |  |    |  |  |  |           |    |  |  |  |  |  |  |  |  |
21      LysLysGluGluProLysLysAlaAlaAlaAlaArgGlyAlaAlaLysArgLysLysPro

37        AlaGlyProArgValSerGluLeuIleGlnLysAlaValSerGlnArgLysGlnArg
            :  |  |  |  |  |  |     |  |     |  |  |  |  |  |  |  |
41      SerSerGlyProSerValSerGluLeuIleGlnLysAlaValThrAlaThrSerArgLysGlnArg

56      SerGlyValSerLeuAlaAlaLeuLysLysAlaLeuAlaAlaGlyGlyTyrAspValGlu
          |  |  |  |  |  |  |  |  |  |  :  |  |  |  |  |  |  |  |  |
61      SerGlyValSerLeuAlaAlaLeuLysLysTyrLeuAlaGlyGlyTyrAspValGlu

76      LysAsnAsnSerArgLeuArgLeuAlaLeuLysSerLeuValSerArgGlnValThrLeuVal
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  :  |  |  |
81      LysAsnAsnSerArgLeuArgLeuAlaLeuLys      ValThrLysGlnThrLeuVal

96      GlnValLysGlyArgValAlaGlyGlnProThrProLysAlaGlyLysGlnAlaGlnThr
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  :
99      GlnValLysGlyArgValAlaArgLysProThrProLysAlaGlyLysGlnAlaGlnAla

116     LysGluLys   ValAlaLysLysLysAlaPro         ValAlaLysLysPro
          |  :  |     |  |  |  |  |  |           |  |  |  |  |
119     LysAspLysAlaAlaAlaLysLysLysAlaProLeuAlaAlaAlaGlnAlaLysLysProAla

131     AlaAlaLysLysProAlaAlaLysLysArgLysLysProLysLysProGlnValAlaAlaLys
          |  |     |    :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
130     AlaAlaAlaLysLysThrAlaLysArgProLysLysProLysLysValAlaAlaAlaAla

151     LysArgProLysLysLysLeuLysLysProGlnAlaLysAlaProAlaLysArgProLysLysPro
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |     |  |  |
159     LysArgProLysLysLeuLysLysProGlnAlaLysAla  AlaLysArgPro

171     LysAlaAlaLysProLysLysValAlaLysArgProAlaLysLysArgValLysProLysLys
          |        |  |    |                          |  |  |  |
175       Ala      LysLysThrAla                       ValLysProLys

191       AlaAlaLysArgProAlaLysAlaLysAlaAlaLysProLysValAlaLysAlaLys
            |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
184     ValAlaAlaLysArgProAlaLysAlaLysAlaAlaLysProLysValAlaLysAlaLys

210     LysAlaAlaProLysLysLysEnd
          |  |  |  |  |  |  |
204     LysAlaAlaProLysLysLysEnd


Matches = 183    Mismatches = 17     Unmatched = 20
Length = 220     Matches/length =  83.2 percent
```

Figure A1a ; Alignment analysis between the X.borealis major cluster histone H1 gene and the X.laevis H1(a) subtype amino acid sequences.
See Table H1 for details.

```
X3302 h3pro1
X3h1 h3pro1

1       MetAlaGlnLeuAlaThrGlnAlaAlaProAlaPheProProProAlaGluPro   AlaGlyLys
        |  |    |  | |  |  |  |  |        | |  | |  |  |    |    |
1       MetThrAlaThrThrGlnThrAlaProValAlaProProAlaGlnProAlaAlaGlyLys

20      Lys   LysLys   GlnProArgLysAlaValGlyAlaGlyLysArgLysProAla
        |     | |    | | | |     |  |  |     | | : | | | :
21      LysThrLysLysGlnLeuProLysLysValAlaLeuGlyAlaGlnLysArgLysProThr

38      GlyProSerValSerGluLeuIleValLysAlaValSerAlaSerLysGlnArgGlnGly
        | |  |     | |  | | |  | |  | |  | :  | |  | |  |  |  |
41      GlyProSerAlaSerGlnLeuIleGlnLysAlaIleSerSerSerLysGlnArgGlyGly

58      ValSerLeuAlaAlaLeuLysLysAlaLeuAlaGlyGlyTyrAspValGluLysAsn
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
61      ValSerLeuAlaLeuLysLysAlaLeuAlaGlyGlyTyrAspValGluLysAsn

78      AsnSerArgLeuIleLeuLeuSerLysValIleSerSerLeuThrLeuValVal
        |  |  |  |  |  |  |  |     | |  :  |    | |    | |
81      AsnSerArgLeuLysLeuLeuLysLysAlaLeuValThrLysGlyTyrThrLeuVal

91      LysGlySerGlyAlaSerGlySerProLysLeuLeuLysLysLysGlnLeuSerThrLysAla
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  :
101     LysGlySerGlyAlaSerGlySerProLysLeuLeuLysLysLysGlnLeuSerThrLysAsp

118     LysValAlaLysLysLys      AlaProValAlaLysLysProAlaAla   LysLys
        |    | |  |  |  |      |  |    | |  |    |  |    |    |
121     LysAlaAlaLysLysLysProAlaAlaProLysAlaLysLysThrAlaAlaGlyAlaLys

139     ProAlaAlaLysAsnArgLysProLysLysValSerGlnAlaAlaLysAsnArgLys
        |     | | | | | | | | | | | | | | | | | | | | | | | | |
141     LysAlaProLysAsnArgLysProLysLysValMetAlaAlaLysAsnArgLys

159     LysLeuLysLysProAlaLysAlaProAlaLysAsnArgProLysLysProLysAlaAlaLys
        |  :  |  |    | | |    |  | |  |  |  |  |  |  |  |  |  |  |
161     LysValLysLysLeuAlaLysAla   AlaLysArgProLysLysProLysAlaValLys

179     ProLysLysValAlaLysLysArgProAlaLysLysArgValLysProLysAlaAlaLysArg
          | |  | |  | |  | |  | |  |  :  |  | |  :  | |  |  |  |
180     AlaLysLysValAlaLysLysArgProAlaLysLysAlaThrLysProLysThrAlaLysArg

199     ProAlaLysAlaLysAlaAlaLysProLysValAlaLysLysAlaLysLysAlaAlaProLys
        |  |  |  |  |    | |  |  |    | |  |  |  |    |  |
200     ProAlaLysAlaLysValAlaLysProLysAlaAlaLysAlaLysLysProAlaPro

219     LysLysEnd
        |  |  |
219     LysLysEnd
```

Matches = 170 Mismatches = 16 Unmatched = 8
Length = 221 Matches/length = 80.1 percent

Figure A3b ; Alignment analysis between the X.borealis major cluster histone H1 gene and the X.laevis H1(b) subtype amino acid sequences.
See Table B1 for details.

```
XB361 hlprot
X3    hlcprot

1       MetAlaGlu  Ala   ThrGlnSerAlaProAlaProProProAlaGluProAlaGly
        | | |    |     | | : : | |     | | | | |
1       MetAlaGluThrAlaSerThrGlnThrThrProAlaAlaProProAlaAlaProLysGln

19      LysLysLysLys   GluProLysLysAlaValGlyGlnAlaLysGlnSerLysLysProAla
        | | | |    | | | | |     |    | | : | | | :
21      LysLysLysLysGlnGlnProLysLysAlaAlaGlyGlyGlnLysAlaLysLysProGlnSer

38      GlyProSerValSerGluLeuIleValLysAlaValSerAlaSerLysGlnArgSerGly
        | | |    | | | | | : | | | | | | |     |
41      GlyProSerAlaSerGluLeuIleValLysSerValSerAlaSerLysGlnArgGlyGly

58      ValSerLeuAlaAlaLeuLysLysAlaLeuAlaAlaGlyGlyTyrAspValAspLysAsn
        | | | | | | | | | | | | | | | | | | | | : : |
61      ValSerLeuAlaAlaLeuLysLysAlaLeuAlaAlaGlyGlyTyrAspValGlnArgAsn

78      AsnSerArgLeuLysLeuAlaLeuLysGlyGlyLeuValSerLysGlnThrLeuValGlnVal
        | | | | | | | | |     | | : |     | |     | |
81      AsnSerArgLeuLysLeuAlaLeuLysAlaLeuValThrLysGlyThrLeuThrGlnVal

98      LysGlySerGlyAlaSerGlySerProAsnSerLeuLysLysGlnLeuAsnGlnLeuThrLysGlu
        | | | | | | | | | | | | | | | | | | | | |     | |
101     LysGlySerGlyAlaSerGlySerProAsnLeuLysLysGlnLeuLeuThrLysGlnVal

118     Lys   ValAlaLysLysLys      AlaProValAlaLysLysPro   AlaAlaLys
        |     | | | | |          | |     | | | |     : | |
121     LysAlaValAlaLysLysLysLeuValAlaProLysAlaLysLysProValThrAlaLys

134     LysProAlaAlaLysSerProLysLysProLysLysValSer   AlaAlaAlaLysSer
        |          | | | | | | | | | |     | | | | |
141     Lys  LysProAspSerProLysLysProLysLysValSerAlaAlaAlaAlaAlaLysSer

153     ProLysLysLeuLysProAlaLysAlaProAlaLysSerProLysLysProLysAla
        | | |    | | |    | |    | | | | | | | | |
160     ProLysLysAlaLysLysProValLysAla   AlaLysSerProLysLysProLysAla

173     AlaLysProLysLysValAlaLysSerProAlaLysLysSerValLysProLysAlaAla
           |    | | | : | | | | | | :     | | | | |
179     ValLysSerLysLysValThrLysSerProAlaLysLysGlnThrLysProLysAlaAla

193     Lys   SerProAlaLysAlaLysAlaAlaLysProLysProValAlaLysAlaLysAlaLys
        |     |          | | | | | | | |     |     | | | |
199     LysAlaLysIleAlaLysAlaLysAlaLysAlaLysGlyLysLysAlaAlaAlaLysLysAsn

213     AlaProLysLysLysEnd


Matches = 172    Mismatches = 37    Unmatched = 17
Length = 226    Matches/length = 76.1 percent
```

Figure A1c : Alignment analysis between the X.boreanlis major cluster histone H1 gene and the X.laevis H1(c) subtype amino acid sequences.
See Table H1 for details.

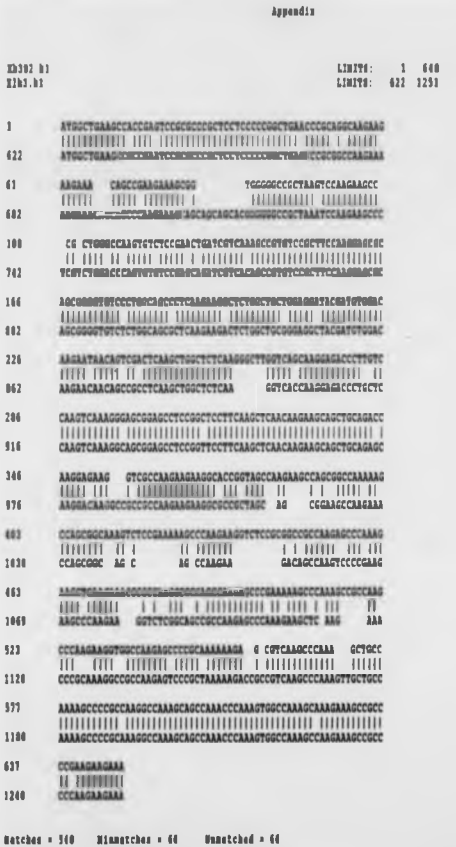


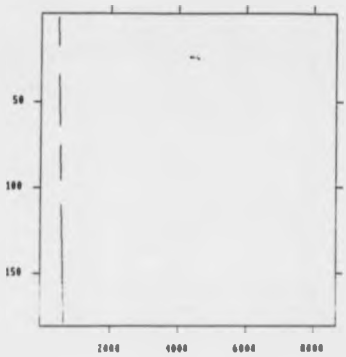
Figure A3d ; Alignment analysis between the X.borealis major cluster histone H1 gene and the X.laevis H1[a] subtype gene sequences.
See Table X1 for details.

```
Xb302.h1                                                LIMITS:    1  648
Xlh1.h1                                                 LIMITS: 10909 11568

1       ATGGCTGAAGCCACCGAGTCCGCGCCCGCTCCTCCCCCGGCTGAAC  CCGCAGGCAAG
        ||| | |   | || ||  |||| || |  ||||||||||| ||||   |||| | ||||
10909   ATGACAGCTACAACTGAAACCGCTCCTGTCGCTCCCCCGGCAGAACCAGCCGCTGCCAAG

58            AAGAAGAAACAGCCGAAGAAAGCGGTGGGGGCCGCTAAGTCCAAGAAGCCCGCT
              |||||| | ||||| |||||||  | ||| | ||| ||  ||||||| ||| |
10969   AAGACAAAGAAGCAGCAGCCTAAGAAAGTAGCGGGAGGCGCAAAAGCCAAGAAACCCTCC

112     GGGCCAAGTGTCTCCGAACTGATCGTCAAAGCCGTGTCCGCTTCCAAGGAGCGCAGCGGG
        ||||| || |  || |||||||||||  |||||||||||| | || ||||||||||||||
11029   GGGCCCAGCGCATCTGAACTGATCGTTAAAGCCGTGTCCTCCTCTAAGGAGCGCAGCGGG

172     GTGTCCCTGGCAGCCCTCAAGAAGGCTCTGGCTGCTGGAGGATACGATGTGGACAAGAAT
        |||||||||||| || ||||||||||||||||||||| ||||||||||||||||||||||
11089   GTGTCCCTGGCCGCTCTCAAGAAGGCTCTGGCTGCCGGAGGATACGATGTGGACAAGAAC

232     AACAGTCGACTCAAGCTGGCTCTCAAGGGCTTGGTCAGCAAGGAGACCCTTGTCCAAGTC
        ||||| || ||||||||||||||||||| ||||||||  |||| ||| ||  |||||||
11149   AACAGCCGCCTCAAGCTGGCTCTCAAGGCCTTGGTCACTAAGGGGACTCTGACCCAAGTC

292     AAAGGGAGCGGAGCCTCCGGCTCCTTCAAGCTCAACAAGAAGCAGCTGCAGACCAAGGAG
        ||||||||||||||||||| ||||||||||||||||||||||||| ||| ||||||||||
11209   AAAGGGAGCGGAGCCTCCGGATCCTTCAAGCTCAACAAGAAGCAACTGGAGACCAAGGAC

352     AAGGTCGCCAAGAAGAAGGCACCGG      TAGCCAAGAAGCCAGCGGCC   AAAAAG
        |||| ||||||||||| || |||        |||||||| | || |||     ||||
11269   AAGGCGGCCAAGAAGAAGCCAGCGGCGCCCAAAGCCAAGAAAACCGCAGCCGGGGCAAAG

403     CCAGCGGCAAAGTCTCCGAAAAAGCCCAAGAAGGTCTCCGCGGCCGCCAAGAGCCCAAAG
          ||| | ||||| || ||||||||||||||||||| || || || |||||||| |||
11329   AAGGCGCCCAAGTCCCCTAAAAAGCCCAAGAAGGTCTCGGCAGCAGCAAAGAGCCCCAG

463     AAGCTCAAGAAACCCGCCAAGGCGCCAGCCAAGAGCCCGAAAAAGCCCAAAGCCGCCAAG
        ||| | |||||||  || |||  ||| ||||| ||||| || || || || | |  ||
11389   AAGGTGAAGAAACTGGCAAAG  GCC GCCAAAAGCCCCAAGAAACCGAAGGCTGTTAAA

523     CCCAAGAAGGTGGCCAAGAGCCCCGCAAAAAAGAGCGTCAAGCCCAAAGCTGCCAAAAGC
        |||||||||||||||||| ||||| ||||||  |  |||||||||| |||||| |||
11446   GCCAAGAAGGTGGCCAAGAGTCCCGCTAAAAAGGCCACCAAGCCCAAAACTGCCAAGAGC

583     CCCGCCAAGGCCAAAGCAGCCAAACCCAAAGTGGCCAAAGCAAAGAAAGCCGCCCCGAAG
        || || ||||||||||  |||||||||||| ||| ||||| ||||  | || || |||
11506   CCAGCAAAGGCCAAAGTCGCCAAACCCAAAGCGGCTAAAGCCAAGAAGCCTGCGCCTAAG

643     AAGAAA
        |||
11566   AAG

Matches = 525    Mismatches = 117    Unmatched = 24
Length = 666     Matches/length = 78.8 percent
```

Figure A3e ; Alignment analysis between the *X.borealis* major cluster histone H1 gene and the *X.laevis* H1(b) subtype gene sequences.
See table H1 for details.

```
Xb302.h1                                                LIMITS:    1  648
Xl h1c

1       ATGGCTGAA  GCC  ACCGAGTCCGCGCCCGGTCCTCCCCCGGCTGAACCC  GCAG
        |||||||||  |||  || ||| |  | |||||  |||||||||| ||||||  |||
1       ATGGCTGAAACTGCCTCAACTGAGACAACTCCCGCCGCTCCCCCGGCAGAACCCAAACAG

53       GCAAGAAGAAGAAACAGCCGAAGAAAGCGGTGGGGGCCGCTAAGTCCAAGAAGCCGCT
         ||||||||| | ||||| ||||| || | ||| | ||||||| ||||||| ||| |
61      AAGAAGAAGAAGCAGCAGCCTAAGAAGGCAGCGGGAGGCGCTAAGGCCAAGAAACCCTCC

112     GGGCCAAGTGTCTCCGAACTGATCGTCAAAGCCGTGTCCGCTTCCAAGGAGCGCAGCGGG
        || || || |  || || |||||||| ||| |||||||||| || ||||||||  | |||
121     GGACCGAGCGCATCTGAGCTGATCGTGAAATCCGTGTCCGCCTCTAAGGAGCGTGGTGGG

172     GTGTCCCTGGCAGCCCTCAAGAAGGCTCTGGCTGCTGGAGGATACGATGTGGACAAGAAT
        |||||||||||| || |||||||||||  ||||||| ||||| |||||||||||| | |||
181     GTGTCCCTGGCCGCTCTCAAGAAGCCCTTGGCTGCCGGAGGTTACGATGTGGAGAGGAAC

232     AACAGTCGACTCAAGCTGGCTCTCAAGGGCTTGGTCAGCAAGGAGACCCTTGTCCAAGTC
        |||||||| |||||| |||||||||||||  |||||||  |||| ||| ||  |||||||
241     AACAGTCGCCTCAAGTTGGCTCTCAAGGCTTGGTCACTAAGGGGACTCTCACCCAAGTC

292     AAAGGGAGCGGAGCCTCCGGCTCCTTCAAGCTCAACAAGAAGCAGCTGCAGACCAAGGAG
        ||||| |||||||||||| || |||||||||||| ||||| |||||||||  ||||||||| |
301     AAAGGCAGCGGAGCCTCTGGATCCTTCAAGCTGAACAAAAAGCAGCTGGAGACCAAGGTG

352     AA  GGTCGCCAAGAAGAA      GGCACCGGTAGCCAAGAAGCCAGCGGCCAAAAAG
        ||  ||| ||||||||||||     ||| ||  |||||||| || |  |  ||||
361     AAGGCGGTGGCCAAGAAGAAGCTCGTGGCGCCCAAAGCCAAGAAACCCGTCACGGCAAAG

403     CCAGCGGCAAAGTCTCCGAAAAAGCCCAAGAAGGTCTCCGCG   GCCGCCAAGAGCCCA
          | | | || || || ||||||||||||||||||||||| |||   || || ||||||||
421     AAAAAGCCCAAATCCCCTAAAAAGCCCAAGAAGGTCTCGGCGGCAGCAGCAAAGAGCCCC

460     AAGAAGCTCAAGAAACCCGCCAAGGCGCCAGCCAAGAGCCCGAAAAAGCCCAAAGCCGCC
        ||||||   |||||||| | |||  ||| ||||| ||||| || ||||||||||| |
481     AAGAAGGCGAAGAAACCGGTAAAG  GCC GCCAAAAGCCCCAAGAAGCCCAAAGCTGTT

520     AAGCCCAAGAAGGTGGCCAAGAGCCCCGCAAAAAAGAGCGTCAAGCCCAAAGCTGCCAAA
        ||  |||||||||||| ||||||| || || |||||| |  ||||||||||||
538     AAATCCAAGAAGGTGACCAAGAGTCCAGCTAAAAAGGCCACTAAGCCCAAAGCT

580     AGCCCCGCCAAGGCCAAAGCAGCCAAACCCAAAGTGGCCAAAGCAAAGAAAGCCGCCCCG
             ||||||||||||  ||||||| |||| | ||| || |  ||||| || ||  |
592          GCCAAGGCCAAAATAGCCAAAGCCAAGGCGGCTAAGGGGAAGAAGGCTGCGGCT

640     AAGAAGAAA
        || |||
646     AAAAAG

Matches = 502    Mismatches = 128    Unmatched = 39
Length = 669     Matches/length =  75.0 percent
```

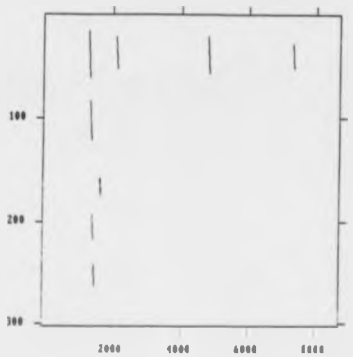
Figure A1f ; Alignment analysis between the X.borealis major cluster histone H1 gene and the X.laevis H1(c) subtype gene sequences. See Table H1 for details.




```
1       AGCTCCCCTAAATGCTCGGTGGGCAAATGAGTGCAGGGGTGGCTACGTTTGTCCAAT
        ||||||||||| || |||   | | |  | |  |||||| ||||| ||||||||||
442     AGCTCCCCTAAAAGCCCGGTCTCGGGCAAGTAAATGAGTGCAGGGGTGGCTACGTTTGTCCAGTT

60      GACCAAATTTTGCAAATGAACGCAGCCAATCAACAAGACGCAACACGTTCTATATA
        || |  | |||| | ||| |||| ||||||||||||| |||| ||||| ||  ||||||
502     AGAACTCAATTTTAC AATAAACTGAGCCAATCAAC AGTCAGAACACCCTT GTATATA

119     AGCAAAAGCGGAAGCGCTAAGGTCAGTTTGTC TTTGTTAA AACGACAGAAATTT
        ||||||| |||||| | ||| |||||| || ||||| || ||||||||||||| |
559     AGTTTAAGTTGTAAAGTCGAAGTCTCGTTTTGTCTTTTGTGAAAAAAACTTCTGTTAATCT

176     GCTA
        || |
619     GCAA
```

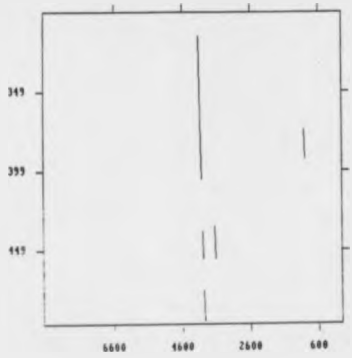
Matches = 139 Mismatches = 37 Unmatched = 8 X3HM102
Length = 184 Matches/Length = 75.5 percent X3h3 L3HT90 442 622

Figure A4.1 A.(Upper) Matrix analysis of X3HM102 intergenic region 1 from figure B3 (ordinate) with the entire X3h3 sequence (abscissa)
B.(lower) Alignment analysis of homologous regions.




```
1        GCACCGTCACTACTAGTGGCCAACTGAACCAAAGGCTCTTTTAAGAGCCACCACATCCCT
         || |  |||||| |||||||||| | |||| ||||||||||||||||||||||||| |||
1266     GCTCGCTCACTA TAGTGGCCAATTCAACC AAGGCTCTTTTAAGAGCCACCACA CCCC

61       CCCGAACGAGCTTCCCCTCTTAACTTGCCCGATCTCCTGCTCTCAACACACGTCTCCT
         || ||| |||||  | |    |||||| |||| ||||||||||||| |||| |||| |||
1323     CCTTGAAAGAGCT TAC    AACTT CCCGTTCTCCTGCTTCTACCACAAGTCTCCCT

121      ACTCAACTCTTTTCTCCCTT
         ||  | |    |  | ||
1376     ACTTGCCCTGAATATTTCTC

Matches = 105    Mismatches = 15 Unmatched = 10    XbmW102  LIN1T9:    1   140
Length = 140     Matches/length =  79.8 percent    Xlh3     LIN1T9: 1266  1395


193      CCAACATGTAGTCGTTTAGCGGTGCCCCTTGCCGC    CCCCAGTGAGGCAGAACCTTC
         || |||||||||||||||| ||||||   ||||    || | |  |    ||| |||
1415     CCCACTTGTTAGTCGTTTAGTGGTTCGGCCGCCCGCGAGTACCAACTCGGCTTAAACATTC

249      TCTTCGGAGGGCGGAG
         | |||||||| ||||
1479     TGTTCGGAGGAGGGAG

Matches = 54     Mismatches = 10     Unmatched = 4   XbmW102  LIN1T9:  193   264
Length = 76      Matches/length =  71.1 percent     Xlh3     LIN1T9: 1415  1490
```

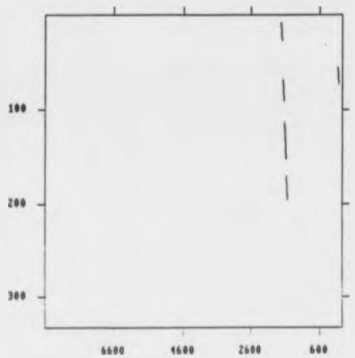
Figure A6.2 A.(Upper) Matrix analysis of XbmW102 intergenic region 2 from figure B1 (ordinate) with the entire Xlh3 sequence (abscissa)
B.(lower) Alignment analysis of homologous regions.




```
313     GAAATATTTAGACAGGTAAACAAACTGGCAGAGTAGCACAGAGCTCAGCGGAGAGGCTCA
        ||||| |||||| | || ||||||| | ||||| | ||| |||||||||| ||||||| ||
4115    GAAATCTTTAGAGAAGT AACAAACGGCCAGAGGAACACCGAGCTCAGCTGAGAGGATCG

373     TTGGCTTGTG CTACTAGTGGGGTCCTGGGGGAAACGGGAAAGACGTAGAAGATAAAGGT
        |||| ||| | ||| |||||||| |  || |     |||||||  ||| | || ||  |
4056    TTTGGTTGCGCCTATTGTGTTTTTC C  GGTG    GGCAAAGGAGTAAAGCAGAACTGG

432     AACGATTTCG TGCAAAGGCTGGACAAAAGAAATA   AATGGCGGAAACCCTGCAGAAA
        |||||||| | |||||||| || || | | |||||   ||   |   || |||||| |||
4004    AACGATTTTGCTGCAAAGGATGTACCAGATAAATAGATAAATACATGAAACCTGCACAAA

488     GTCTATCC
        ||||||||
3944    GTCTATCC
```

```
Matches = 133 Mismatches = 17  Unmatched = 14   XbW102          LIMITS:   313   499
Length = 188  Matches/length =  72.9 percent    Xlh3    PYRAMB & LIMITS:  3937  4115
```

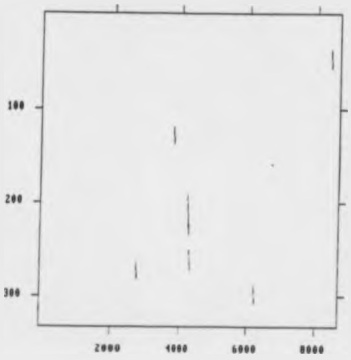
Figure A4.3 A.(Upper) Matrix analysis of XbW0102 intergenic region 3 from figure B3 (ordinate) with the entire Xlh3 sequence (abscissa)
B.(lower) Alignment analysis of homologous regions.




```
1       CATAATTCAAAGACTAGACAAAGTCAAACAGCGAGTTAGCACTAATAAAGCAATACTGCT
        |||  |||||  |||||||| ||| |    | |  |  || ||  | | | |||| |
1704    CAT GTTCAGCAACTAGACAGGATCA   GTCTA  TAAAGCAAACGATGAAGTGCTGCT

61      CCGTCCCTGCAACGGCTTTTATAGCCTGCACCATGCAAAAGAAACTTGGGCTGGTACACT
        |||   ||||||| |||||||| |||| ||||  |||| | || | | |  ||||
1650    CCG  GCTGCAGCTGCTTTTATACCCTGC CCATG TAAAGCA CTGGCCCGTGACACT

121     CCTTCCCATTGGCTGGCTCTGTAGCATGACGTCCCTTTCTTGCGTTTGCGTGAAACCGT
          |||||||||||||||||| ||||||||||       || ||| | |  |||||||
1595    GATTCCCATTGGCTGGCTCTGTAGCATGACGTCCCTTTCTTGCGTTTGCGTGAAACCGT

181     TGGCGGTAAGT AAACTT
        ||| ||||||| ||||||
1537    TGCTGGTAAGTAAAGCTT
```

```
Matches = 141    Mismatches = 43     Unmatched = 14    XbaW362          LIMITS:     1   197
Length = 198     Matches/Length =  71.2 percent       Xba1   pYAAN A   LIMITS:  1520  1704
```

Figure A4.4 A.(Upper) Matrix analysis of XbaW362 intergenic region (from figure B1 (ordinate) with the entire Xba1 sequence (abscissa)
B.(lower) Alignment analysis of homologous regions.

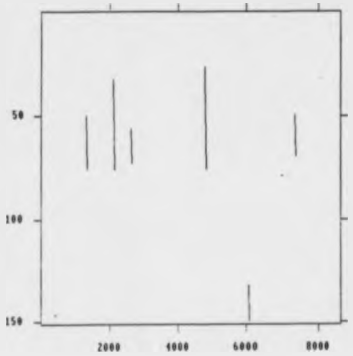



```
192     AAGCTTGCTCTTCCTTTGGCTGCCGCAGCAAGCC TAACCAATCCGGC AGAGAGAGGAGCTCT

252     GGCAGCTGCCTATAAAAAGGCAAAGGC
        || |||| |||||||| | ||   |
4306    GCTGCTTTTTTTATATTTTTTTT
```

Matches = 67 Mismatches = 17 Unmatched = 2 XbW302 LIMITS: 192 257
Length = 86 Matches/Length = 77.9 percent Xlb2 LIMITS: 4240 4331

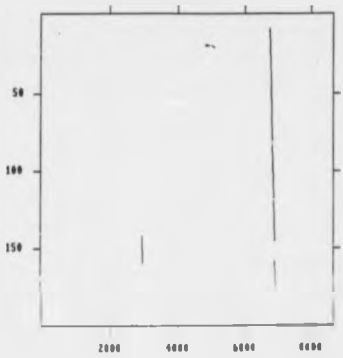
Figure A4.5 A.(Upper) Matrix analysis of XbW302 intergenic region 5 from figure B3 (ordinate) with the entire Xlb2 sequence (abscissa)
B.(lower) Alignment analysis of homologous regions.

172




```
1       TGAGCCTCCACGCACACAACGCAACCATCTGCCCTCACCACACACCCAAACGGCTCTTTT
        |||| ||| | | | |||   || | ||||||| ||||| |||||||| |||||||||
4769    TGAG CTCTTCCCTC CAAAAATACTA CTGCCCTGACCACTCACCCAAA GGCTCTTTT

61      CAGAGCCA CCACTATCTCTCAA AGCGCTGT
        |||||||| |||||   ||| || ||||||||
4825    CAGAGCCACCCACTTCATCTAAACAGCGCTGT

Matches = 71     Mismatches = 15     Unmatched = 6   XhBW302  LIKE79:    1    80
Length = 92      Matches/length =  77.2 percent      Xlh3     LIKE79:  4769  4836
```

Figure A6.6 A.(Upper) Matrix analysis of XhBW302 intergenic region 6 from figure B3 (ordinate) with the entire Xlh3 sequence (abscissa)
B.(lower) Alignment analysis of homologous regions.




```
1       AAGCTTATTTTGAAGAATGCTCCTGGGGGTTTGCACTTTATCCAATCACGAACTGGCCCA
        |||    ||| | ||||||||| |||| || ||||||| ||||||||||||||| | | ||
6764    AAA ACTTTTGGGAATTGCTACTGGGGA TTGACGTTTATCCAATCACGAATTAACACA

61      TAGAACTGCTAGAAGCGCTGTTGGGAAAGGGCGGGGTGGTA GTGCAGTTCCCTATCAGGT
        | ||||| ||||||| ||  | || ||||||||||||||| ||| ||||||||||| ||
6822    TCGAACTCGTAGAAGCCGATG GGAAGGGCGCGGGTGGTACTTCACGTCCCTATCA GT

121     CAGCCCAATGTGTCATCATGCTGTGGGGTGTAAGCGGGTAGGTCACATAGTCGAGAGCTA
        |||||||||||||||||| | | || | |  |||| | |||||||| | |||| |   |
6880    CAGCCCAATGTGTCATATGA TTTAGGGTCTGAGCGCCTGAACTCACACATTCGATAGGA

181     GCTTCGAGATTTAAGT
          || | || | || |
6939    CTTTAGTGAGTGAAAT
```

```
Matches = 150    Mismatches = 41    Unmatched = 5  XhH0302  LIMITS:    1  196
Length = 196     Matches/length =  76.5 percent    XlhI    LIMITS:  6764  6954
```

Figure A4.7 A.(Upper) Matrix analysis of XhH0302 intergenic region 7 from figure B3 (ordinate) with the entire XlhI sequence (abscissa)
B.(lower) Alignment analysis of homologous regions.

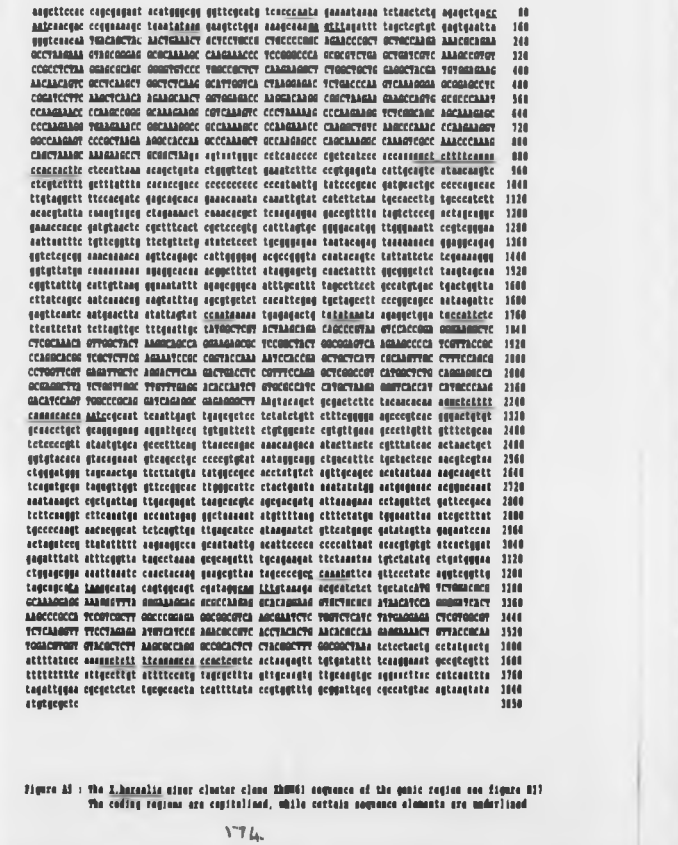


Figure A3 : The X.hermalia minor cluster clone XHM61 sequence of the genic region see figure B17
The coding regions are capitalised, while certain sequence elements are underlined

```
Start                                                                  199
ATG ACA GCT ACA ACT GAA ACT GCT CCT GCC GCT GCC CCG GCA GAA CCC GCT GCT GCC AAG
Met Thr Ala Thr Thr Glu Thr Ala Pro Ala Ala Ala Pro Ala Glu Pro Ala Ala Ala Lys

                                                                       259
AAA ACG CAG AAG CCT AAG AAA GTA GCG GGA GGC GCA AAA GCC AAG AAA CCC TCC GGG CCC
Lys Thr Gln Lys Pro Lys Lys Val Ala Gly Gly Ala Lys Ala Lys Lys Pro Ser Gly Pro

                                                                       319
AGC GCG TCT GAG CTG ATC GTC AAA GCC GTG TCC GCC TCT AAG GAG CGC AGC GGG GTG TCC
Ser Ala Ser Glu Leu Ile Val Lys Ala Val Ser Ala Ser Lys Glu Arg Ser Gly Val Ser

                                                                       379
CTG GCC GCT CTC AAG AAG GCT CTG GCT GCT GGA GGC TAC GAT GTG GAG AAG AAC AAC AGT
Leu Ala Ala Leu Lys Lys Ala Leu Ala Ala Gly Gly Tyr Asp Val Glu Lys Asn Asn Ser

                                                                       439
CGC CTC AAG CTG GCT CTC AAG GCA TTG GTC ACT AAG GAG ACT CTG ACC CAA GTC AAA GGG
Arg Leu Lys Leu Ala Leu Lys Ala Leu Val Thr Lys Glu Thr Leu Thr Gln Val Lys Gly

                                                                       499
AGC GGA GCC TCC GGA TCC TTC AAG CTC AAC AAG AAG CAA CTG GTG GAG ACC AAG GAC AAG
Ser Gly Ala Ser Gly Ser Phe Lys Leu Asn Lys Lys Gln Leu Val Glu Thr Lys Asp Lys

                                                                       559
GCG GCT AAG AAG AAG CCA GTG GCG CCC AAA TCC AAG AAA CCC CAA GCC GGG GCA AAG AAG
Ala Ala Lys Lys Lys Pro Val Ala Pro Lys Ser Lys Lys Pro Gln Ala Gly Ala Lys Lys

                                                                       619
GCG TCA AAG TCC CCT AAA AAG CCC AAG AAG GTC TCG GCA GCA GCA AAG AGC CCC AAG AAG
Ala Ser Lys Ser Pro Lys Lys Pro Lys Lys Val Ser Ala Ala Ala Lys Ser Pro Lys Lys

                                                                       679
GTG AAG AAA CCG GCA AAG GCC GCC AAA AGC CCC AAG AAA CCC AAG GCT GTC AAG CCC AAA
Val Lys Lys Pro Ala Lys Ala Ala Lys Ser Pro Lys Lys Pro Lys Ala Val Lys Pro Lys

                                                                       739
CCC AAG AAG GTG GCC AAG AGT CCC GCT AAG AAG GCC ACC AAG CCC AAA GCT GCC AAG AGC
Pro Lys Lys Val Ala Lys Ser Pro Ala Lys Lys Ala Thr Lys Pro Lys Ala Ala Lys Ser

                                                                       799
CCA GCA AAG GCC AAA GTC GCC AAA CCC AAA GCA GCT AAA GCA AAG AAG CCT GCG GCT AAG
Pro Ala Lys Ala Lys Val Ala Lys Pro Lys Ala Ala Lys Ala Lys Lys Pro Ala Ala Lys

AAG TAA
Lys End
```

Figure A6(a) : Translation of Minor Cluster XbHW51 H1 gene.

```
Start                                1021                                1051
ATG GCT CGT ACT AAG CAG ACA GCC CGT AAG TCC ACC GGA GGG AAG GCT CCT CGC AAA CAG
Met Ala Arg Thr Lys Gln Thr Ala Arg Lys Ser Thr Gly Gly Lys Ala Pro Arg Lys Gln
                                1081                                1111
TTG GCT ACT AAG GCA GCC AGG AAG AGC GCT CCG GCT ACT GGC GGA GTC AAG AAG CCC CAT
Leu Ala Thr Lys Ala Ala Arg Lys Ser Ala Pro Ala Thr Gly Gly Val Lys Lys Pro His
                                1141                                1171
CGT TAC CGC CCA GGC ACG GTC GCT CTT CGA GAA ATC CGC CGC TAC CAA AAA TCC ACC GAG
Arg Tyr Arg Pro Gly Thr Val Ala Leu Arg Glu Ile Arg Arg Tyr Gln Lys Ser Thr Glu
                                1201                                1231
CTG CTC ATT CGC AAG TTG CCT TTC CAG CGC CTG GTT CGT GAG ATT GCT CAG GAC TTC AAG
Leu Leu Ile Arg Lys Leu Pro Phe Gln Arg Leu Val Arg Glu Ile Ala Gln Asp Phe Lys
                                1261                                1291
ACT GAC CTC CGT TTC CAG AGC TCG GCC GTC ATG GCT CTG CAG GAG GCC AGC GAG GCT TAT
Thr Asp Leu Arg Phe Gln Ser Ser Ala Val Met Ala Leu Gln Glu Ala Ser Glu Ala Tyr
                                1321                                1351
CTG GTT GGC TTG TTT GAG GAC ACC AAT CTG TGC GCC ATC CAT GCT AAG AGG GTC ACC ATC
Leu Val Gly Leu Phe Glu Asp Thr Asn Leu Cys Ala Ile His Ala Lys Arg Val Thr Ile
                                1381
ATG CCC AAG GAC ATC CAG TTG GCC CGC AGG ATC AGA GGC GAG AGG GCT TAA
Met Pro Lys Asp Ile Gln Leu Ala Arg Arg Ile Arg Gly Glu Arg Ala End
```

Figure A6(b) : Translation of Minor Cluster XhUW61 H3 gene.

```
Start                                1297                                1327
ATG TCT GGA CGC GGC AAA GGA GGA AAG GGT TTA GGC AAA GGA GGC GCC AAG AGC CAC AGC
Met Ser Gly Arg Gly Lys Gly Gly Lys Gly Leu Gly Lys Gly Gly Ala Lys Arg His Arg
                                1357                                1387
AAG GTG CTG CGC CAT AAC ATC CAG GGG ATC ACT AAG CCC GCC ATC CGT CGC TTG GCC CGC
Lys Val Leu Arg Asp Asn Ile Gln Gly Ile Thr Lys Pro Ala Ile Arg Arg Leu Ala Arg
                                1417                                1447
AGA GGC GGC GTC AAG CGA ATC TCT GGT CTC ATC TAT GAG GAG ACT CGT GGC GTT CTC AAG
Arg Gly Gly Val Lys Arg Ile Ser Gly Leu Ile Tyr Glu Glu Thr Arg Gly Val Leu Lys
                                1477                                1507
GTT TTC CTG GAG AAT GTC ATC CGA GAC GCC GTC ACC TAC ACT GAA CAC GCC AAG AGG AAA
Val Phe Leu Glu Asn Val Ile Arg Asp Ala Val Thr Tyr Thr Glu His Ala Lys Arg Lys
                                1537                                1567
ACT GTT ACC GCA ATG GAC GTG GTG TAC GCT CTT AAG CGC CAG GGC CGC ACT CTC TAC GGC
Thr Val Thr Ala Met Asp Val Val Tyr Ala Leu Lys Arg Gln Gly Arg Thr Leu Tyr Gly

TTT GGC GGC TAA
Phe Gly Gly End
```

Figure A6(c) : Translation of Minor Cluster XhUW61 H4 gene.

```
XhH1 hiprot
Xlh3.hiaprot

1       MetThrAlaSerThrGluThrAlaProAlaAlaAlaProAlaAlaGluProAlaAlaLysAla
          |  :    :  :  |  :  |  |       | |  |  |  |  |     |
1       MetAlaGlyAlaAlaGluSerAlaProAlaProProProAlaGluProAlaAlaLysLys

21      LysThrGlnLysProLysLysValAla         GlyGlyAlaLysAlaLysLysPro
          |    |    | |  |    (              |    | |  :  | |  |
21      LysLysGlnGlnProLysLysAlaAlaAlaAlaArgGlyAlaAlaLysSerLysLysPro

38        SerGlySerSerAlaSerGluLeuIleGlnLysAlaValSerAlaSerLysGlnArg
            | |  |    | |    | |    | |  |  |  |  |  |  |  |  |
41      SerSerGlyProSerValSerGluIleIleGlnThrAlaValSerAlaSerLysGlnArg

57      SerGlyValSerLeuAlaAlaLeuLysLysAlaLeuAlaAlaGlyGlyTyrAspValGlu
          | | | | | | | | | | : | | | | | | | | | :
61      SerGlyValSerLeuAlaAlaLeuLysLysThrLeuAlaAlaGlyGlyTyrAspValAsp

77      LysAsnAsnSerArgLeuTyrLeuAlaLeuLysLeuThrValThrLysGluThrLeuGlu
          | | | | | | | | | | |       | | | | | |
81      LysAsnAsnSerArgLeuTyrLeuAlaLeuLys      ValThrLysGluThrLeuLeu

97      GlnValLysGlySerGlyAlaSerGlySerPheTyrLeuAsnLysTyrGlnLeuValGlu
          | | | | | | | | | | | | | | | | | | | |
99      GlnValLysGlySerGlyAlaAspGlySerPheTyrLeuAsnLysTyrGlnAlaLeu  Glu

117     ThrTyrAspTyr   AlaAlaTyrTyrLys   ProValAlaProLysSerLysTyrPro
          : | | |    | | | | |    | : |       : | | |
118     SerTyrAspTyrAlaAlaAlaTyrTyrLysAlaProLysAspIleAlaIleLysTyrPro

135     GlnAlaGlyGlnIleGlnLeuSerIleProSerProLysProLysTyrGlnSerAlaLeu
          |   | | | : : | | | | | | | | | | | | | | | | |
136     AlaAlaAlaLeuLysIleThrAlaTyrAspProLysProLysProLysTyrGlnSerAlaLeu

153     AlaTyrAspProLysIleGlyLeuGlyProAlaLeuMetAlaTyrAspProLysTyrPro
          | | | | | | : | | | | | | | | | | | |    |
150     AlaIleSerProLysIleLeuMetIleProAlaLeuAlaLeuSerProAlaLysIle

175     LysAlaValTyrProTyrProLysTyrValAlaLeuSerProAlaLysLysAlaThrTyr
          | |       | | |       | | | | |     | |
170     ThrAlaVal      LysProTyrValAlaLeuSerProAla   LysAla

193     ProLysAlaLeuSerProAlaLysAlaLysValAlaLysProLysAlaLeuAlaLysAla
          | | | |    |       | |   | |     | | |
183       LysAlaLeuLys   ProLysValAlaLys   AlaLys   LysAlaLeuProLys

213     LysLysProAlaLeuAlaLysLysEnd
          | |
200     LysLysEnd

Matches = 169    Mismatches = 40     Unmatched = 13
Length = 228     Matches/Length =  72.4 percent
```

Figure A7(a) : Comparison of X.laevis clone XlH1 histone H1A sequence [Perry et al 1985] with the X.borealis oiner cluster clone XbOM61.H1 sequence.

```
Xh61h1prot
xlh1prot

1       MetThrAlaThrThrGlnThrAlaProAlaAlaAlaProAlaMetAlaProAlaAlaAlaLys
          |  |  |  |  |  |  |  |  |  |     |  |  |  |  |  |  |  |  |
1       MetThrAlaThrThrGlnThrAlaProValAlaProProAlaMetAlaProAlaAlaAlaLys

21      LysThr      GlnLysProLysLysValAlaGlyGlnAlaLysAlaAlaLysLysProGlnGlr
          |  |      |    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
21      LysThrLysLysGlnAlaProLysLysValAlaAlaGlyGlnAlaLysAlaLysLysProGlnGlr

39      GlyProGlnAlaAsnGlnLeuLeuValLysAlaValSerAlaSerLysGlnArgGlnGlu
          |  |  |  |  |  |  |  |  |  |  |  |  |  :  |  |  |  |  |
41      GlyProGlnAlaGlnGlnLeuLeuValLysAlaValGlnAlaSerLysGlnArgGlnGlu

59      ValSerLeuAlaAlaLeuLysLysAlaLeuAlaAlaGlyGlyTyrAspValGluLysAsn
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  :  |  |
61      ValSerLeuAlaAlaLeuLysLysAlaLeuAlaAlaGlyGlyTyrAspValGluAspAsn

79      AsnSerArgLeuLysLeuAlaLeuSerAlaAlaAsnValThrLysSerThrGlnThrVal
          |  |  |  |  |  |  |  |  |  |  |  |  |     |  |  |  |  |
81      AsnSerArgLeuLysLeuAlaLeuLysAlaAlaValThrLysGlyThrLeuThrGlnVal

99      LysGlySerGlyAlaSerGlySerThrLysLeuAsnLysLysProGlnLeuGlnValGlyThrLys
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
101     LysGlySerGlyAlaSerGlyGlyThrLysLeuAsnAsnLysLysGlnLeuGlnValGlyThrLys

119     AspLysAlaAlaLysLysLysProValAlaProLysSerLysLysProGlnAlaAlaGlyAla
          |  |  |  |  |  |  |  |  |  |  |  :  |  |  |  |  |  |  |  |
120     AspLysAlaAlaLysLysLysProLysProAlaProLysAlaLysLysProThrAlaAlaGlyAla

139     LysLysAlaSerLysAlaSerProLysLysProLysLysValSerAlaAlaAlaLysAlaSerPro
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
140     LysLysAlaProLysAlaSerProLysLysProLysLysValSerAlaAlaAlaLysAlaSerPro

159     LysLysValLysLysProAlaAlaAlaLysAlaSerProLysLysProLysAlaAlaValLys
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
160     LysLysValLysLysProLysAlaAlaLysAlaSerProLysLysProLysAlaAlaVal

179     ProLysProLysLysValAlaLysLysSerProAlaLysLysLysAlaValThrLysProLysAlaAla
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
179       LysAlaLysLysValAlaLysLysSerProAlaLysLysAlaValThrLysProThrAla

199     LysSerProAlaLysAlaLysLysValAlaLysProLysAlaAlaLysAlaLysLysProAlaLys
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
198     LysSerProAlaLysAlaLysLysValAlaLysProLysAlaAlaLysAlaLysLysProAlaLys

219     AlaLysLysEnd
          |  |  |
218     ProLysLysEnd

Matches = 204     Mismatches = 15     Unmatched = 5
Length = 224      Matches/length = 91.1 percent
```

Figure A7(b) : Comparison of X.laevis clone X1h1 histone H10 sequence [Perry et al 1985] with the X.borealis minor cluster clone Xh0961.H1 sequence.

```
Xb41.hlprot
Xl.hlcprot

1       Met      ThrAlaThrThrGlnThrAlaProAlaAlaAlaProAlaAlaProAlaAla
                  |     | | : | | | : | | |    | | | |
1       MetAlaAlaValThrAlaAsnThrGlnThrThrProAlaAlaAlaProProAlaAlaProAlaGln

19      AlaLysLysThrGlnLysProLysGlnValAlaAlaGlyGlyAlaArgAlaLysLysProAsn
            | |     |    | | |    | | | | | | | | | |
21      LysLysLysLysGlnAlaProLysLysAlaAlaGlyGlyAlaLysAlaLysLysProAsn

39      GlyProAsnAlaAsnGluLeuIleValLysAlaValThrAlaAsnLysGlnArgAsnGly
          | | | | | | | | | | : | | | | | | |    |
41      GlyProAsnAlaAsnGluLeuIleValLysAsnAsnValAsnAlaAsnLysGlnArgAsnGly

59      ValSerLeuAlaAlaLeuLysLysAlaLeuAlaAlaGlyGlyProArgProValAlaLysAsn
          | | | | | | | | | | | | | | | | | | : |
61      ValSerLeuAlaAlaLeuLysLysAlaLeuAlaAlaGlyGlyProArgProValGlnAsnAsn

79      AsnAsnArgAsnLysLeuAlaLeuLysAlaAsnValThrLysGlnThrAsnThrGlnVal
          | | | | | | | | | | | | |    | | | | |
81      AsnAsnArgAsnLysLeuAlaLeuLysAlaAsnValThrLysGlyThrAsnThrGlnVal

99      LysGlyAsnGlyAlaAsnGlyAsnThrAsnAsnAsnLysAsnLeuAsnValAlaThrIle
          | | | | | | | | | | | | | | | | | | |
101     LysGlyAsnGlyAlaAsnGlyAsnThrAsnAsnAsnLysAsnGlnAsn  GlnThrIle

119     AspLysAla  AlaLysLysIleProValAlaProLysAsnLysLysProGlnAlaGly
            | |    | | | |    | | | | : | | |
120     ValLysAlaValAlaLysLysLysLeuValAlaProLysAlaLysLysProPro  ValThr

138     AlaLysLysAlaAsnLysAsnProLysLysProLysLysLysValAsn  AlaAlaAlaLys
          | | |       | | | | | | | | | | |    | | | |
139     AlaLysLysLysProLysAsnProProLysLysProLysLysValAsnAlaAlaAlaLys

157     AsnProLysLysValLysLysProAlaLysLysAlaAlaAlaLysAsnProProLysLysProLysAla
          | | | |    | | |    | | | | | | | | | | | | | | | |
159     AsnProLysLysAlaLysLysProProLysLysAlaAlaAlaLysAsnProLysLysProLysAla

177     ValLysProLysProLysLysValAlaLysLysAsnProAlaLysLysAlaValThrLysProLys
          |       |    | | | : | | | | | | | | | | | |
179     Val     LysAsnLysLysValThrLysLysAsnProAlaLysLysAlaValThrLysProLys

197     AlaAlaLysAsnProAlaLysAlaLysValAlaLysProLysAlaAlaLysAlaLysAlaLys
          |             | | | | : | |    | | | |    | |
197     Ala           AlaLysAlaAlaLysAlaLysAlaAlaLysAlaAlaLysAlaLysAlaLys

217     ProAlaAlaLysLysAsn
            | | | | |
213     AlaAlaAlaLysLysMet

Matches = 186    Mismatches = 28    Unmatched = 12
Length = 226    Matches/length = 82.3 percent
```

Figure A7(c) : Comparison of X.laevis clone X12 histone H1C sequence [Turner et al 1983] with the X.borealis ovary cluster clone Xb0041.01 sequence.

```
Xb61.h1                                                    LIMITS:   170   832
Xlb3.h1                                                    LIMITS:   622  1251

170     ATGACAGCTACAACTGAAACTGCTCCTGCCGCTGCCCCGGCAGAACCCGCTGCTGCCAAG
        ||| | |   |  | ||| | || || ||  || |||||||| || ||||| ||    ||
622     ATGGCTGAAGCCGCCGAATCCGCCCCCGCTCCTCCCCCGGCTGAGCCCGCGGCCAAGAAA

230     AAAACGCAGAAGCCTAAGAAAGTAG        CGGGAGG CGCAAAAGCCAAGAAACCC
        || |  ||| |||| |||||||| ||        |||| || ||| ||| |||||||| |||
682     AAGAAACAGCAGCCCAAGAAAGCAGCAGCAGCACGGGGGGCCGCTAAATCCAAGAAGCCC

281       TCCGGGCCCAGCGCGTCTGAGCTGATCGTCAAAGCCGTGTCCGCCTCTAAGGAGCGC
          || || ||||| | ||| |||| ||||||||| |||||||||||| || |||||||||
742     TCGTCTGGACCCAGTGTGTCCGAGCAGATCGTCACAGCCGTGTCCGCTTCCAAGGAGCGC

338     AGCGGGGTGTCCCTGGCCGCTCTCAAGAAGGCTCTGGCTGCTGGAGGCTACGATGTGGAG
        ||||||||||| ||||| || ||||||||| |||||||||| ||||||||||||||||
802     AGCGGGGTGTCTCTGGCAGCGCTCAAGAAGACTCTGGCTGCGGGAGGCTACGATGTGGAC

398     AAGAACAACAGTCGCCTCAAGCTGGCTCTCAAGGCATTGGTCACTAAGGAGACTCTGACC
        |||||||||||| ||||||||||||||||||||      |||||| |||||||| ||| |
862     AAGAACAACAGCCGCCTCAAGCTGGCTCTCAA      GGTCACCAAGGAGACCCTGCTC

458     CAAGTCAAAGGGAGCGGAGCCTCCGGATCCTTCAAGCTCAACAAGAAGCAACTGGTGGAG
        |||||||||| |||||||||||| ||||||||||||||||||||||||  | || ||
916     CAAGTCAAAGGCAGCGGAGCCTCCGGTTCCTTCAAGCTCAACAAGAAGCA  GCTGCAG

518     ACCAAGGACAAG   GCGGCTAAGAAGAA   GCCAGTGGCGCCCAAATCCAAGAAACCC
        | ||||||||||   || || ||||||||   |||  | ||  |  || ||||||||||
973     AGCAAGGACAAGGCCGCCGCCAAGAAGAAGGCGCCCGCTAGCAGCGGAAGCCAAGAAACCA

572     CAAGCCGGGGCAAAGAAGGCGTCAAAGTCCCCTAAAAAGCCCAAGAAGGTCTCGGCAGCA
          || |  || |||||| |  | ||||||||| || ||||||||||||||||||||||||
1033    GCGGCAGCAGCCAAGAAGACAGCCAAGTCCCCGAAGAAGCCCAAGAAGGTCTCGGCAGCC

632     GCAAAGAGCCCCAAGAAGGTGAAGAAACCGGCAAAGGCCGCCAAAAGCCCCAAGAAACCC
        || |||||||| |||||| | |||||||| ||||||||||||||| || |||   |||
1093    GCCAAGAGCCCAAAGAAGCTCAAGAAACCCGCAAAGGCCGCCAAGAGTCCCGCTAAAAAG

692     AAGGCTGTCAAGCCCAAACCCAAGAAGGTGGCCAAGAGTCCCGCTAAG   AAGGCCACC
        | || ||||||||||     |||  | | ||||| || ||||| |||   || ||  ||
1153    ACCGCCGTCAAGCCCA     AAGTTGCT GCCAAAAGCCCCGCAAAGGCCAAAGCAGCC

749     AAGCCCAAAGCTGCCAAGAGCCCAGCAAAGGCCAAAGTCGCCAAACCCAAAGCAGCTAAA
        || |||||||  ||||| |||| || ||| |||   | | ||||    |||
1207    AAACCCAAAGTGGCCAA AGCCAG AAA GCC   GCCCCCAAGAAGAAA

809     GCAAAGAAGCCTGCGGCTAAGAAG


Matches = 491    Mismatches = 118    Unmatched = 75
Length = 684     Matches/length =  71.8 percent
```

Figure A7(d) : Comparison of X.laevis clone Xlh1 histone H1A gene sequence [Perry et al 1985] with the X.borealis minor cluster clone XbHW61.H1 gene sequence.

Xb61.h1 LIMITS: 170 832
Xlh1.h1 LIMITS: 10909 11568

```
170     ATGACAGCTACAACTGAAACTGCTCCTGCCGCTGCCCCGGCAGAACCCGCTGCTGCCAAG
        ||||||||||||||||||| ||||||| |||| ||||||||||||| || |||||||||
10909   ATGACAGCTACAACTGAAACCGCTCCTGTCGCTCCCCCGGCAGAACCAGCCGCTGCCAAG

230     AAAAC    GCAGAAGCCTAAGAAAGTAGCGGGAGGCGCAAAAGCCAAGAAACCCTCC
        || ||    |||| ||||||||||||||||||||||||||||||||||||||||||||
10969   AAGACAAAGAAGCAGCAGCCTAAGAAAGTAGCGGGAGGCGCAAAAGCCAAGAAACCCTCC

284     GGGCCCAGCGCGTCTGAGCTGATCGTCAAAGCCGTGTCCGCCTCTAAGGAGCGCAGCGGG
        |||||||||| ||||| |||||||| |||||||||| ||||||||||||||||||||||
11029   GGGCCCAGCGCATCTGAACTGATCGTTAAAGCCGTGTCCTCCTCTAAGGAGCGCAGCGGG

344     GTGTCCCTGGCCGCTCTCAAGAAGGCTCTGGCTGCTGGAGGCTACGATGTGGAGAAGAAC
        |||||||||||||||||||||||||||||||||||| ||||| |||||||||| ||||||
11089   GTGTCCCTGGCCGCTCTCAAGAAGGCTCTGGCTGCCGGAGGATACGATGTGGACAAGAAC

404     AACAGTCGCCTCAAGCTGGCTCTCAAGGCATTGGTCACTAAGGAGACTCTGACCCAAGTC
        ||||| |||||||||||||||||||| |||||||||||| ||||||||||||||||||
11149   AACAGCCGCCTCAAGCTGGCTCTCAAGGCCTTGGTCACTAAGGGGACTCTGACCCAAGTC

464     AAAGGGAGCGGAGCCTCCGGATCCTTCAAGCTCAACAAGAAGCAACTGGTGGAGACCAAG
        ||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||
11209   AAAGGGAGCGGAGCCTCCGGATCCTTCAAGCTCAACAAGAAGCAACT TTGGAGCCAAG

524     GACAAGGCGGCTAAGAAGAAGCCAGTGGCGCCCAAATCCAAGAAACCCCAAGCCGGGGCA
        |||||||||| |||||||||||| |||||||||| |||||||| || ||||||||||
11266   GACAAGGCGGCCAAGAAGAAGCCAGCGGCGCCCAAAGCCAAGAAAACCGCAGCCGGGGCA

584     AAGAAGGCGTCAAAGTCCCCTAAAAAAGCCCAAGAAGGTCTCGGCAGCAGCAAAGAGCCCC
        ||||||||| | ||||||||||||||||||||||||||||||||||||||||||||||||
11326   AAGAAGGCGCCCAAGTCCCCTAAAAAAGCCCAAGAAGGTCTCGGCAGCAGCAAAGAGCCCC

644     AAGAAGGTGAAGAAACCGGCAAAGGCCGCCAAAAGCCCCAAGAAACCCAAGGCTGTCAAG
        |||||||||||||||| |||||||||||||||||||||||||||| ||||||||
11386   AAGAAGGTGAAGAAACTGGCAAAGGCCGCCAAAAGCCCCAAGAAACCGAAGGCTGT

704     CCCAAACCCAAGAAGGTGGCCAAGAGTCCCGCTAAGAAGGCCACCAAGCCCAAAGCTGCC
          ||| |||||||||||||||||||||||||||||| |||||||||||||||| |||||
11442     TAAAGCCAAGAAGGTGGCCAAGAGTCCCGCTAAAAAGGCCACCAAGCCCAAAACTGCC

764     AAGAGCCCAGCAAAGGCCAAAGTCGCCAAACCCAAAGCAGCTAAAGCAAAGAAGCCTGCG
        |||||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||
11509   AAGAGCCCAGCAAAGGCCAAAGTCGCCAAACCCAAAGCGGCTAAAGCCAAGAAGCCTGCG

824     GCTAAGAAG
        |||||||||
11568   CCTAAGAAG

Matches = 630    Mismatches = 34     Unmatched = 19
Length = 683     Matches/length =  92.7 percent
```

Figure A5(a) : Comparison of X.laevis clone Xlh1 histone H1B gene sequence (Perry et al 1988) with the X.borealis oliver cluster clone Xb0061.H1 gene sequence.

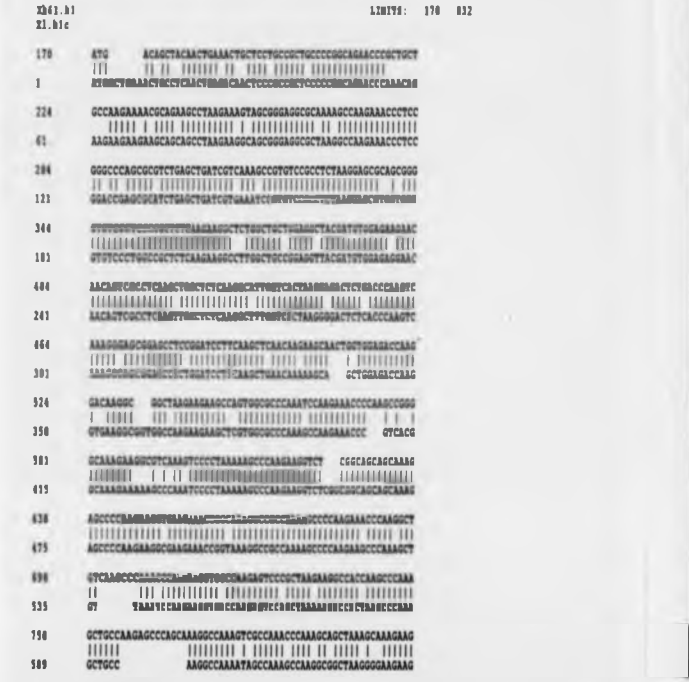
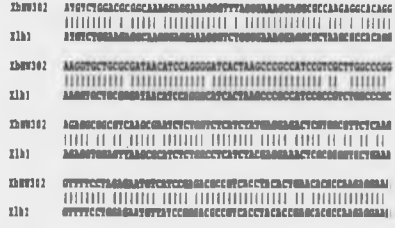
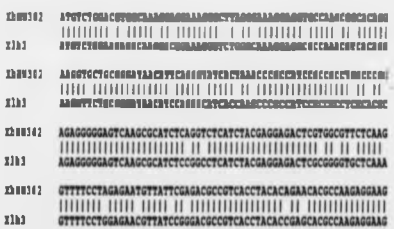


Figure A7(f) : Comparison of X.laevis clone X1h1 histone H1C gene sequence [Turner et al 1983] with the X.borealis minor cluster clone XbH61.H1 gene sequence.



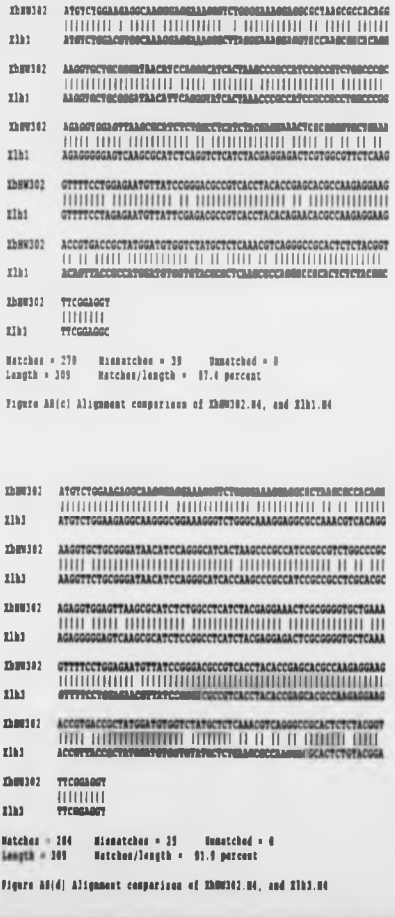
Matches = 244 Mismatches = 45 Unmatched = 0
Length = 309 Matches/length = 85.4 percent

Figure A8(a) Alignment comparison of XbMV41.R4, and XbMV382.R4



Matches = 268 Mismatches = 41 Unmatched = 0
Length = 309 Matches/length = 86.7 percent

Figure A8(b) Alignment comparison of Xlb1.R4, and Xlb3.R4

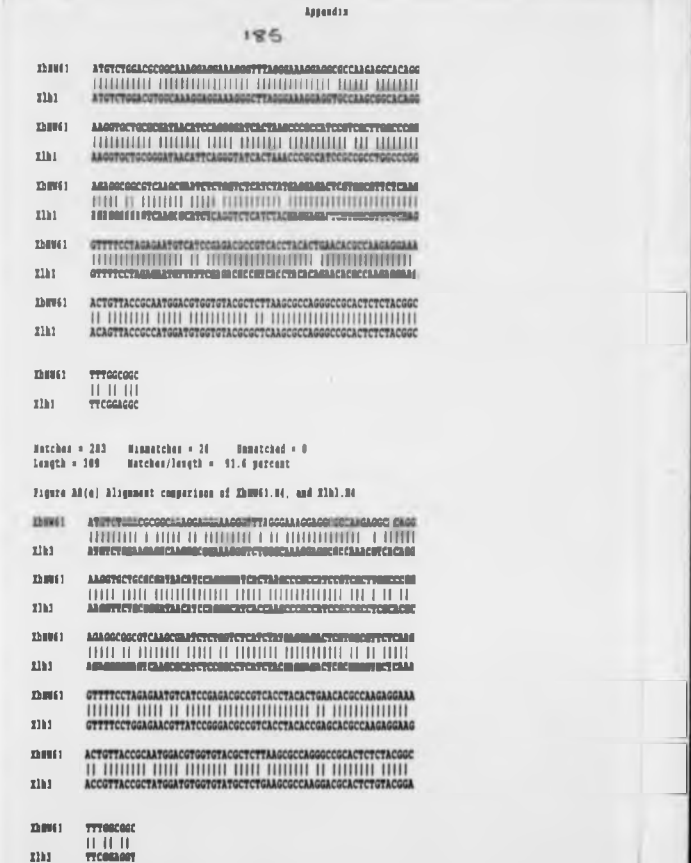


Matches = 270 Mismatches = 39 Unmatched = 0
Length = 309 Matches/length = 87.4 percent

Figure A8(c) Alignment comparison of XbHW302.R4, and Xlb1.R4

Matches = 284 Mismatches = 25 Unmatched = 0
Length = 309 Matches/length = 91.9 percent

Figure A8(d) Alignment comparison of XbHW302.R4, and Xlb3.R4



Matches = 283 Mismatches = 26 Unmatched = 0
Length = 309 Matches/length = 91.6 percent

Figure A8(e) Alignment comparison of Xbm961.R4, and Xlb1.R4

Matches = 281 Mismatches = 48 Unmatched = 0
Length = 309 Matches/length = 84.5 percent

Figure A8(f) Alignment comparison of Xbm961.R4, and Xlb1.R4

REFERENCES

Bagshaw,J.C., Andrews,H.V., Posy,H.A., (1984) Organization and expression of histone genes in the brine shrimp Artemia. In 'Histone genes and Histone gene expression, ed. G.Stein, J.Stein, N.Marzluff. New York: Wiley

Bailey,G., Cocks,G., Wilson,A., (1969) Gene duplication in fishes: malate dehydrogenase of salmon and trout. Biochem. Biophy. Res. Commun 34 605-612

Baltimore,D. (1981) Gene conversion: Some implications for Immunoglobulin genes. Cell 24 592-594

Barralle,F.,E., and Brownlee,G.C., (1978) AUG is the only recognizable signal sequence in the 5' non-coding regions of eukaryotic mRNA. Nature 274 84-87

Benoist,C., O'Hare,K.O., Breathnach,R., Chambon,P., (1980) The ovalbumin gene: Sequence of putative control regions. Nucl.Acids Res. 8 127-142

Benton,W.D., Davis,R.W., (1977) Screening Lambda GT recombinant clones by hybridization to single plaques in situ. Science 196 180-182

Birchmeier,C., Grosschedl,R., Birnstiel, M.L. (1982) Generation of authentic 3' termini of an H2A mRNA in vivo is dependent on a short inverted DNA repeat and on spacer sequences. Cell 28 739-745

Birchmeier,C., Schumperli,D., Sconzo,G., Birnstiel, M.L. (1984) 3' editing of mRNAs: Sequence requirements and involvement of a 60 nucleotide DNA in maturation of histone mRNA precursors. Proc.Nat.Acad.Sci.USA 81 1057-1061

Bird,A., Taggart,M., Macleod,D., (1981) Loss of rDNA methylation accompanies the onset of ribosomal gene activity in early development of X. laevis. Cell 26 381-390

Birnstiel,M.L., Fulford,J., Wainberg,M.S., Stafford,D., (1974) Isolation and some properties of the genes coding for histone proteins. Proc.Nat.Acad.Sci.USA 71 2900-2904

Birnstiel,M.L., Kressmann,A., Schaffner,W., Portmann,R., Busslinger,M. (1978). Phil. Trans. Royal Soc. (London) B 283 319-324

Birnstiel,M.L., Busslinger,M. Strub,K. (1985) Transcription Termination and 3' processing: The End is in Site. Cell 41 349-359

References

Birnstiel,M.L., Portmann,R., Busslinger,M. Schaffner,W., Probst,E., Kressmann,A., (1979) Alfred Benzon Symp 13 117-129.

Bishop,C.A., Maher,M.A., Wilson,A.C., Hadji-Azimia,I. Fischberg,M., (1977) Albumen phylogeny for clawed frogs. Science 195 785-87

Brandt,W.F., Strickland,W.N., Strickland,M., Carlisle,D., Woods,D., von Holt,C. (1979) A histone programme during the life cycle of the sea urchin. Eur. J. Biochem. 94 1-10

Brandt,W.F., Patterson,E., von Holt,C., (1980) The histones of yeast: The isolation and partial structure of the core histones. Eur. J. Biochem. 110 67-76

Breathnach,R., Chambon,P., (1981) Organisation and expression of eukaryotic split genes coding for proteins. Ann. Rev. Biochem 50 349-83

Busslinger,M. Portmann,R, Birnstiel,M. (1979) A regulatory sequence near the 3' end of sea urchin histone genes. Nucleic Acids Res. 6 2997-3008

Busslinger,M. Portmann,R, Irminger,J., Birnstiel,M. (1980) Ubiquitous and gene specific regulatory 5' sequences in a sea urchin histone DNA clone coding for histone protein variants. Nucleic Acids Res. 8 957-977

Busslinger,M. Rusconi,S., Birnstiel,M.L. (1982) An unusual evolutionary behavior of a sea urchin histone gene cluster. EMBO Journal 1 27-33

Chandler,M.E., Kedes,L.H., Cohn,R.H., Yunis,J.J. (1979) Genes coding for histone proteins in Man are located on the distal end of the long arm of chromosome 7. Science 205 908-910

Childs,G., Maxson,R., Kedes,L.H., (1979) Histone gene expression during sea urchin embryogenesis: Isolation and characterization of early and late messenger RNAs of S. purpuratus by gene specific hybridization and template activity. Dev.Biol. 73 153-173

Childs,G., Maxson,R., Cohn,R., Kedes,L.H., (1981) Orphons: Dispersed genetic elements derived from tandemly repetitive genes of eukaryotes. Cell 23 651-661

Childs G., Nocente-McGrath,C., Lieber,T., Holt,C., Knoulen,J., (1982) Sea Urchin (Lytechinus pictus) late stage histone H3 and H4 genes: Characterization and mapping of a clustered but non-tandemly repeated linked multigene family. Cell 31 383-93

Chae,J., Kalodrubetz,D., Crumstain,N. (1982) The two yeast histone H2A genes encode similar protein subtypes. Proc.Nat.Acad.Sci.USA 79 1484-87

Clewell,D.B., Helinski,D.R., (1970) Properties of a supercoiled deoxyribonucleic acid-protein relaxation complex and strand specificity of the relaxation event. Biochemistry 9 4428-4440

Clewell,D.B. (1972) Nature of Col E plasmid replication in E. coli in the presence of chloramphenicol. J. Bact. 110, 667-676

Coen,E., Strachan,T., Dover,G. (1982) Dynamics of concerted evolution of ribosomal DNA and histone gene families in the melanogaster species of the subgroup of Drosophila. J. Mol. Biol. 158 17-35

Cohen,L.H., Newrock,K.H., Zweidler,A. (1975) Stage specific switches in histone synthesis during embryogenesis of the sea urchin. Science 190 994-997

Cohen,L.H., Newrock,K.H., Hendricks,M.B., Donelly,R.J., Weinberg,H.S. (1978) Stage specific mRNAs coding for subtypes of H2A and H2B histones in the sea urchin embryo. Cell 14 327-336

Cohn,R.H., Kedes,L.H. (1979a) Nonallelic histone gene clusters of individual sea urchins (Lytechinus pictus): Mapping of homologies in coding & spacer DNA. Cell 18 855-64

Cohn,R.H., Kedes,L.H. (1979b) Nonallelic histone gene clusters of individual Sea Urchins (Lytechinus pictus): Polarity and gene organization. Cell 18 843-54

Comings,D., Mosquita,J., Hirbein,R.J. Staten,J.C. Biron,C.H., (1984) Organization of the histone genes in the Rainbow Trout (Salmo gairdneri). J. Mol. Evol. 20 227-235

Cordon,J., Chalylyk,B., Buchwalder,A., Sonnenn-Corni,P. Kedringer,C., Chambon,P., (1980) Promotor sequences of eukaryotic protein coding genes. Science 209 1406-1414

Crawford,E.J., Krieg,P., Harvey,R.P., Hewith,D.A., Wells,J.R.E. (1979) Histone genes are clustered with a 15 Kb repeat in the chicken genome. Nature 279 133-136

D'Andrea,R., Harvey,R.P., Wells,J.R.E. (1981) Vertebrate histone genes: Nucleotide sequence of a chicken H2A gene and regulatory flanking sequences. Nucl.Acids Res. 9 1119-1136.

Darnell,J.E., (1982) Variety in the level of gene control in eukaryotic cells. Nature 297 365-371

Delange,R.J., Fambrough,D.M., Smith,E., Bonner,J., (1969) The complete amino acid sequence of calf thymus histone H4: presence of N-acetyllysine. J.Biol. Chem. 244 319-334

Dalmaso,E.J., Williams,L.C., Searcy,D.G., (1981) A histone-like protein (HTa) from Thermoplasma acidophilum. J.Biol. Chem. 256 905-911

Dentran,O.H.J., Bandiq,M.M., De Leaf,E.T.M., Kemtar,J.G. (1984) Organisation of Xenopus histone gene variants within clusters and their transcriptional expression. Biochim. Biophys. Acta 740 390-401

Diaz,M.O., Barsacchi-Pilone, Mahon,K.A., Gall,J.G. (1981) Transcripts from both strands of a satellite DNA occur on lampbrush chromosome loops of the Newt Notomphthalmus. Cell 24 649-659

Durn,A.J., Hallenbeck,S., Straub,A., Benoist,C., Mathis,D. (1987) A multiplicity of CCAAt box-binding protein Cell 50 863-872

Dover,G.A (1982) Molecular Drive: A cohesive mode of species evolution. Nature 299 111-117

Dover,G.A (1986) Molecular drive in multigene families:how biological novelties arise spread and are assimilated. TIG 2 (6) 159-165

Dover,G.A., Brown,S.D.M., Coen,E.S. Dallas,J. Strachan,T., Trick,M. (1982) In 'Genome Evolution' (eds. Dover,G.A. & Flavell,R.B.) pp343-374 Academic press, London

Efstradiatis,A., Posakony,J.W., Maniatis,T., Lawn,R.M., O'Connell,C., Spritz,R.A., De Riel,J.K., Forget,B.G, Weissmann,S.M., Slightom,J.L., Blechl,A.E., Smithies,O., Baralle,F.E., Shoulders,C.C., Proudfoot,N.J. (1980) The Structure and evolution of the human-globin gene family. Cell 21 653-668

Elsevier,S.M., Lippe,M.J. Steinbeck,C. (1978) Histone genes in macronuclear DNA of the ciliate Stylonychia mytilus. Chromosoma 69 291-301

Engel,J.D., Dodgson,J.B. (1981) Histone genes are clustered but not tandemly repeated in the chicken genome. Proc.Nat.Acad.Sci.USA 78, 2856-2860

Engel,J.D., Sugarman,B., Dodgson,J.B., (1982) A chicken histone H3 gene contains intervening sequences. Nature 297, 434-36

Engel J.D., (1984) Organisation and expression of chicken histone genes. In 'Histone genes and Histone gene expression, ed. G.Stein, J.Stein, W.Marzluff. New York: Wiley

Federoff,N.V., (1979) On spacers. Cell 16 697-710

References

Fletcher,C. Heintz,N., Roeder, G., (1988) Purification and characterization of OTF-1 a transcription factor regulating cell cycle expression of a human histone H2b gene. Cell 51 773-781

Frischauf,A., Lehrach,H., Poustka,A. & Murray,N. (1983) Lambda replacement vectors carrying polylinker sequences. J.Mol.Biol. 170 827-842

Fritzgerald,M., Shenk,T., (1981) The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. Cell 24 251-260

Galli,G., Hofstetter,H., Stunnenberg,H.G., Birnstiel,M.L. (1983) Biochemical complimentation with RNA in the Xenopus oocyte: A small RNA is required for the generation of 3' histone mRNA termini. Cell 34 823-828

Gallinari,P., La Bella,F., Heintz,N. (1988) Characterization and purification of H1TF2, a novel CCAAT-binding protein that interacts with a Histone H1 subtype-specific consensus element. Mol. Cell. Biol 8 1566-1575

Gimlitis,A.A., Stankevicicte,J.V., Vorobev,V.I. (1976) Chromatin proteins from normal, vegetalized and animalized sea urchin embryos. Dev. Biol. 52 181-192

Goldberg,M.L. (1979) PhD Thesis, Stanford University

Grunstin,A.S., (1981) Two temporal phases for the control of histone gene activity in cleaving sea urchin embryos Dev. Biol. 87 163-175

Grunstin,A.S., Hilt,F.H. (1981) Protein synthesis, polyribosomes and peptide elongation in early development of S. purpuratus. Dev. Biol. 82 32-40

Grandy,D.K., Engel,J.D., Dodgson,J.B. (1982) Complete nucleotide sequence of a chicken H2B gene. J.Biol.Chem 257 8577-80.

Grunz,H.U., Schaffner,W., Telford,J., Birnstiel,M.L., (1976) Molecular analysis of the histone gene cluster of Psammechinus miliaris: II The arrangement of the five histone coding and spacer sequences. Cell 9 471-478

Gronchadl,R., Birnstiel,M.L. (1988a) Spacer DNA sequences upstream of the 5'-TATAAATA-3' sequences are essential for promotion of H2A histone gene transcription. Proc.Natl.Acad.Sci.USA 77 7102-06

References

Grosschedl,R., Birnstiel,M.L. (1980b) Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. Proc.Natl.Acad.Sci.USA 77 1432-36

Grosschedl,R., Birnstiel,M.L. (1982) Delimitation of far upstream sequences required for maximal in vitro transcription of an H2A histone gene. Proc.Natl.Acad.Sci.USA 77 7102-06

Grosveld,G.C., Shoemaker,C.K., Jat,P., Flavell,R.A., (1981) Localization of DNA sequences necessary for transcription of rabbit globin gene in vitro. Cell 25 215-226

Grunstein,M., Schedl,P., Kedes,L.H., (1976) Sequence analysis and evolution of sea urchin (Lytechinus pictus and Strongylocentrotus purpuratus) histone H4 messenger RNAs. J. Mol. Biol. 104 351-369

Grunstein ,M. (1978) Hatching in the sea urchin L. pictus is accompanied by a shift in histone H4 gene activity. Proc.Nat.Acad.Sci.USA 75 4135-4139

Grunstein,M., Diamond,K.H., Kreppel,H., Grunstein,J., (1981) Comparison of the early histone H4 gene sequence of Strongylocentrotus purpuratus with maternal, early and late histone H4 mRNA sequences. Biochem. 20 1216-1223

Harvey,R.P., Kreiq,P.A., Robins,A.J., Wells,J.R.E., (1981) Non-tandem arrangement and divergent transcription of chicken histone genes. Nature 294 49-53

Harvey,R.P., Robins,A.J., Wells,J.R.E.,(1982) Independently evolving chicken H2B genes: Identification of a ubiquitous H2B-specific 5' element. Nucl.Acids Res. 10 7851-7863

Harvey,R.P., Whiting,J.A., Colen,L.S., Kreiq,P.A., Wells,J.R.E. (1983) H2Af4 An extremely variant histone H2A sequence expressed in the chick embryo. Proc. Natl. Acad. Sci.USA 80 2819-23

Harvey,R.P., Wells,J.R.E. (1984) Chicken histone genes and their variants. In 'Histone genes and Histone gene expression, ed. G.Stein, J.Stein, N.Marzluff. New York: Wiley

Heintz,N., Sarnit,H., Roeder,R.G., (1981) The structure of the human histone genes: Clustered but not tandemly repeated. Cell 24 661-668.

Hentschel,C.C., Birnstiel,M.L., (1981) The organization and expression of histone gene families. Cell 25 301-313

Busteschel,C.C., Probst,E., Birnstiel,M.L., (1980a) Transcriptional fidelity of histone genes injected into Xenopus oocyte nuclei. Nature 284 100-102

Busteschel,C.C., Irminger,J.C., Busser,P., Birnstiel,M.L., (1980b) Sea urchin histone mRNA termini are located in gene regions downstream from putative regulatory sequences. Nature 285 147-151

Busteschel,C.C. (1982) Homocopolymer sequences in the spacer of sea urchin histone gene repeat are sensitive to S1 nuclease. Nature 295 714-716

Bureford,L., Fahrer,K., Woodford,J., Rosbash,M., (1979) Isolation of yeast histone genes H2A and H2B. Cell 18 1261-1271.

Bieter,P., Hendricks,M., Hemminki,K., Weinberg,K., (1979) Histone gene switch in the sea urchin embryo. Identification of late embryonic histone messenger RNAs and the control of their synthesis. Biochemistry 18 2702-2716

Bilder,V.A., Livermay,B.M., Turner,P.C., Vlad,M.P. (1981) Histone gene number in relation to C-value in amphibians. Nucl.Acids Res. 9 5737-5745

Binegardner,R.T. (1967) Echinoderms. In Methods in Developmental Biology pp139-55 New York: Crowell

Bo,S.L., Manley,J.L., (1981) DNA sequences required for initiation of transcription in vitro from the major late promotor of adenovirus 2. Proc.Nat.Acad.Sci.USA 78 830-834

Bohacher,U., Lutz,H., Hornberg,A., (1980) Novel histone H2A-like protein of H. COLI. Proc.Nat.Acad.Sci.USA 77 5097-5101

Bort,M.H., Pandey,N.B., Marzluff,W.F., (1984) A region of the coding sequence is required for high level expression of a murine histone H3 gene. Proc.Nat.Acad.Sci.USA 84 4450-4454

Jacob,E. (1976) Histone gene reiteration in the genome of the mouse. Eur. J. Biochem 65 275-284

Jacob,E., Malacinski,G.A., Birnstiel,M.L. (1976) Reiteration Frequency of the histone genes in the genome of the Amphibian Xenopus laevis. Eur. J. Biochem. 68 45-54

Jeffreys,A.J. (1982) Recent studies of gene evolution using recombinant DNA in 'Genome Evolution' (eds. Dover,G.A. & Flavell,R.B.) pp1-48 Academic press, London

John,H.A., Birnstiel,M.L., Jones,K.W., (1969) RNA-DNA hybrids at the cytological level. Nature 223 582-587

References

Kedes,L.H., Birnstiel,M.L., (1971) Reiteration frequency and clustering of DNA sequences complimentary to histone messenger RNAs. Nature 230 165-170

Kedes,L.H., Chang,A.C.Y., Houseman,D. and Cohen,S.N. (1975a) Isolation of histone genes from unfractionated sea urchin DNA by subculture cloning in E. coli. Nature 255, 533-538

Kedes,L.H., Cohn,R.H., Lowry,J.C., Chang,A.C.Y., Cohen,S.N. (1975b) The organization of sea urchin histone genes. Cell 6 359-369

Kedes,L.H. (1979) Histone genes and histone messengers. Ann.Rev.Biochem. 48 837-870.

Kedes,L.H., Crane,P.D., Cognetti,G., Hunter,A.L. (1969) Synthesis of nuclear and chromosomal proteins on light polyribosomes during cleavage in the sea urchin embryo. J. Mol. Biol. 45 337-351

Kimura,M. (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature (Lond) 267 275-276

Klein,H.L., Petes,T.D. (1981) Intrachromosomal gene conversion in yeast. Nature 289 144-148

Knochel,W., Karge,E., Wanner,A., Meyerhof,W., (1986) Globin evolution in the genus Xenopus: Comparative analysis of cDNAs coding for adult globin polypeptides of X. borealis and X. tropicalis. J. mol. Biol. 23 211-233

Kornberg,R.D., (1977) Structure of Chromatin. Ann. Rev. Biochem. 46 931-954

Kossel,A. (1884) Cited by H. Bush in "Histones and Other Proteins". (1965) Academic Press; New York

Kreig,P.A., Robins,A.J., Colman,A., Wells,J.R.E., (1982) Chicken histone H5 mRNA: the polyadenylated RNA lacks the conserved histone 3' terminator sequence. Nucl.Acids Res. 10 6777-6785

Kreig,P.A., Robins,A.J., D'Andrea,R., Wells,J.R.E., (1983) The chicken H5 is unlinked to core and H1 histone genes. Nucl.Acids Res. 11 619-627

Lathe,R., Bur,H., Lecocq,J.P., Heuts,E.K.F. (1980) Prokaryotic histone like protein interacting with RNA polymerase. Proc.Nat.Acad.Sci.USA 77 3548-3563

Levy,S., Sures,I., Kedes,L.H. (1978) Sequence of the 5'-end of S. purpuratus H2B histone mRNA and its location within histone DNA. Nature 276 737-39

Levy,S., Sures,I., Kedes,L.H. (1982) The Nucleotide and Amino Acid Sequence of a gene for H1 Histone that interacts with Euchromatin. J. Biol. Chem. 257 9430-9443

References

Lifton,R.P., Goldberg,M.L., Karp,R.W., Hogness,D.S., (1978) The organization of the histone genes in Drosophila melanogaster: Functional and evolutionary implications. Cold Spring Harbour Symp. Quant. Biol. 42 1047-1051

Lillie,M.L., Green,M.R., (1989) Activator's target in sight. Nature 341 279-280

Lipps,H.J., Steinbruck,G. (1978) Free genes for rRNAs in the macronuclear genome of the ciliate Stylonychia mytilis. Chromosoma 69 21-66

Loenen,W.A. Brammar,W.J., (1980) A bacterial Lambda vector for cloning large DNA fragments made with several restriction enzymes. Gene 10 249-259

Long,E.O., Dawid,I.B. (1980) Repeated genes in eukaryotes. Ann.Rev. Biochem. 49 727-764

Maniatis,T., Fritsch,E.F., Sambrook,J., (1982) Molecular Cloning: A laboratory manual. Cold Spring Harbour Laboratory. New York.

Mardian,J.K.W., Isenberg,I., (1978) Yeast inner histones and evolutionary conservation of histone-histone interactions. Biochem. 17 3425-3433

Mercluff,H.F., Graves,R.A., (1984) Organization and expression of mouse histone genes. In 'Histone genes and Histone gene expression, ed. G.Stein, J.Stein, H.Mercluff. New York: Wiley

Maxson,R., Mahon,K.,Cornaccano,C, Childs,G., Kedes,L.H. (1983a) Distinct organizations and patterns of expression of early and late histone gene sets in the sea urchin. Nature 301 120-125

Maxson,R., Cohn,R., Kedes,L.H., (1983b) Histone genes. In 'Eukaryotic Genes: Their Structure, Activity, and Regulation'. MacLean,N., Gregory,S., Flavell,R., ed. London; Butterworths

Maxson,R., Cohn,R., Kedes,L.H., Mohun,K., (1983c) Expression and Organization of histone genes. Ann.Rev.Genet. 17 239-277

McGhee,J.D., Felsenfeld,G., (1980) Nucleosome structure. Ann. Rev. Biochem. 49 1115-1156

Messing,J., Crea,R., Seeburg,P.H., (1981) A system for shotgun DNA sequencing. Nucl.Acids Res. 9 309-321

Messing,J., (1983) New M13 vectors for cloning. Methods Enzymol. 101 20-79

Montell,C., Fisher,E.F., Caruthers,M.H., Berk,A.J. (1983) Inhibition of RNA cleavage but not polyadenylation by a point mutation in U' consensus sequence AAUAAA. Nature 305 600-605

References

Moorman,A.F.M., De Laaf,R.T.H., Dentres,C.E.J., Telford,J., Birnstiel M.L. (1980) Histone genes from Xenopus laevis: Molecular cloning and initial characterization. Gene 10 185-191

Moorman,A.F.M., De Moor,P.A.J., De Laaf,R.T.H., Van Denegen,H.M.A.M., Dentres,C.E.J. (1980) Primary structure of the histone H3 and H4 genes and their flanking sequences in a minor histone cluster of Xenopus laevis. Febs letters 136 45-52

Murray,N.E., Brammar,W.J., Murray,K., (1977) Lambdoid Phages that simplify the recovery of in vitro recombinants. Mol. Gen. Genet. 150 53-61

Murray,N.E., Murray,K. (1974) Manipulation of restriction targets in phage lambda to form receptor chromosomes for DNA fragments. Nature 251 476

Newrock,K.M., Cohen,L., Hendricks,M., Donnelly,R., Weinberg,R. (1978) Stage specific mRNAs coding from subtypes of H2A and H2B histones in the sea urchin embryo. Cell 14 327-336

Nordheim,N. Rich,A. (1983) Negatively supercoiled simian virus 40 DNA contains Z-DNA segments with transcriptional enhancer sequences. Nature 303 674-679

Ohno,S., (1970) Evolution by gene duplication. Springer-Verlag. New York pp101-105

Old,R.W., Woodland,H.R., Ballentine,J.E.M., Aldridge,T.C., Newton,C.A., Heins,H.A., Turner,P.C. (1982) Organization and expression of cloned histone gene clusters from Xenopus laevis and Xenopus borealis. Nucl.Acids Res. 10 7561-7580

Old,R.W., Woodland,H.R., (1984) Histone genes: Not so simple after all. Cell 38 625-627

Onley, M.A. Gould,J., Kim,H., Kane,M., Haruford,L. (1986) Identification of sequences in a yeast histone promoter involved in periodic transcription. Cell 45 537-544

Overton,G.C., Weinberg,R.S. (1978) Length and sequence heterogeneity of histone gene repeat unit of the sea urchin S. Purpuratus. Cell 14 247-257

Penyin,S., Hilab,B., Chalkay,R., (1971) An electrophoretic comparison of vertebrate histones. J.Biol. Chem. 246 4206-4215

Perchs, H.L., Kedes,L.H., Weinberg,R.S. Birnstiel,M.L. (1977) Localization of sequences coding for histone messenger RNA in the chromosomes of Drosophila melanogaster. Chromosoma 63 135-151

References

Perry,R. Thomas,C.H., Bondar,R.G. (1985) Genomic organization and nucleotide sequences of two distinct histone gene clusters from Xenopus laevis. J. Mol. Biol. 185 479-499

Petes,T.D., (1980) Unequal meiotic recombination within tandem arrays of yeast ribosomal DNA genes. Cell 19 765-74

Portmann,R., Shaffner,H., Birnstiel,M. (1976) Partial denaturation mapping of cloned DNA from sea urchin Psammechinus miliaris. Nature 264 31-34

Prihmen,D., (1979) In "Biological Regulation and Development 1, ed. Goldberg,R.F. New York; Plenum press 250-277

Radloff,R., Bauer,W., Vinograd,J., (1967) A dye-buoyant density method for the detection and isolation of closed circular duplex DNA: The closed circular DNA in HeLa cells. Proc.Nat.Acad.Sci.USA 57 1514-1521

Roberts,S.B., Ramiola,E., Ramsay,S.W., Childs,G., (1987) Molecular characterization of the histone gene family of Caenorhabditis elegans. J. Mol. Biol 196 27-38

Roberts,S.B., Ramsay,S.W., Childs,G., (1989) Nucleotide sequence of Caenorhabditis elegans core histone genes. J. Mol. Biol 206 567-577

Roberti,I., Fragapane,P., Piersandrei-Amaldi,P., Buccuri,R., Amaldi,F., Beccari,I., (1982) Characterization of histone genes isolated from Xenopus laevis and Xenopus tropicalis genomic libraries. Nucl.Acids. Res. 10 7543-59

Saigo,K., Hillestein,L., Thomas Jr.,C.A., (1980) The organization of Drosophila melanogaster histone genes. Cold Spring Harbor Symp. Quant. Biol. 45 815-827

Sanger,F., Coulson,A.R., Hong,G.F., Hill,.D.F., Petersen,B. (1982) Nucleotide sequence of bacteriophage lambda DNA. J. Mol. Biol. 162 729-773

Schaffner,W., Kunz,G., Daneyler, H., Telford,J., Smith,H.O., Birnstiel,M.L. (1978) Genes and spacers of cloned sea urchin histone DNA analyzed by sequencing. Cell 14 655-671

Scherer,S., Davis,R., (1980) Recombination of dispersed repeated sequences in yeast. Science 209 1380-1384

Schumperli,D. (1986) Multilevel regulation of replication dependent histone genes. TIG 4 187-191

References

Seale,R.L., Aronson,A.I., (1973) Chromatin associated proteins of the developing sea urchin embryo 2: Acid soluble proteins. J. Mol. Biol. 75 647-658

Seiler-Tuyns,A., Birnstiel M.L. (1981) Structure and expression in L-cells of a cloned H4 histone gene of the mouse. J.Mol.Biol. 151 607-625.

Sierra,F., Lichtler,A., Marashi,F., Rickles,R., Van Dyke,T., Clark,S., Wells,J., Stein,G., Stein,J. (1982) Organization of human histone genes. Proc.Nat.Acad.Sci.USA 79 1795-1799.

Sittman,D.B., Chui,I., Pun,C., Cohn,R.H., Kedes,L.H. Marzluff,W.F. (1981) Isolation of two clusters of mouse histone genes. Proc.Nat.Acad.Sci.USA 78 4078-4082.

Sittman,D.B., Graves,R.A., Marzluff,W.F. (1981) Structure of a cluster of mouse histone genes. Nucl.Acids Res. 12 6679-6697

Smith,G.P., (1976) Evolution of repeated DNA sequences by unequal cross-over. Science 191 528-535

Smith,M.M., Murray,K. (1983) Yeast H3 and H4 histone messenger RNAs are transcribed from two non-allelic gene sets. J. Mol. Biol. 169 641-661

Smith,M.M. (1984) The organization of the yeast histone genes. In 'Histone genes and Histone gene expression, ed. G.Stein, J.Stein, W.Marzluff. New York: Wiley

Sommer,A., (1978) Yeast chromatin: Search for histone H1. Mol. Gen. Genet 161 323-331

Sommerville,J. (1979) Gene Expression in Lampbrush Chromosomes. Proc. FEBS Meet. 51 265-276 1978

Southern,E.M., (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J.Mol. Biol. 98 503-517

Spradling,A.C. Rubin,G.M. (1981) Drosophila genome organization: Conserved and dynamic aspects. Ann. Rev Genet 15 219-264

Stein,G.S., Sierra,F., Stein,J.L., Plumb,M.,Marashi,F., Carozzi,N., Prokopp,K., Baumbach,L. (1984) Organization and expression of human histone genes. In 'Histone genes and Histone gene expression, ed. G.Stein, J.Stein, W.Marzluff. New York: Wiley

Steinemann,M., (1982) Analysis of chromosomal homologies between two species of the subgroup melanogaster: D. miranda and D. melanogaster using cloned DNA segments. Chromosoma 87 7-88

References

Stephanson,N.C., Erba,H.P., Gall,J.G. (1981a) Characterization of a histone gene cluster of the newt Notophthalmus viridescens. Nucl.Acids Res. 9 2281-295

Stephanson,N.C., Erba,H.P., Gall,J.G. (1981b) Histone gene clusters of the newt Notophthalmus are separated by long tracts of satellite DNA. Cell 24 639-647

Stephanson,N.C., (1984) Organization and expression of newt histone genes. In 'Histone genes and Histone gene expression, ed. G.Stein, J.Stein, N.Marzluff. New York: Wiley

Strub,K., Galli,G., Busslinger,M., Birnstiel, M.L. (1984) The sequences of the sea urchin U7 small nuclear DNA suggest specific contacts between histone mRNA precursor and U7 RNA during RNA processing. EMBO J. 3 2801-2807

Stunnenberg,H.G., Birnstiel.M.L. (1982) Bioassay for components regulating eukaryote gene expression: A chromosomal factor involved in the generation of histone mRNA 3' termini. Proc.Nat.Acad.Sci.USA 79 6201-6204

Sugarman,B.J., Dodgson,J., Engel,J.D. (1983) Genomic organization, DNA sequence and expression of chicken embryonic histine genes. J.Biol.Chem 258 9005-9016

Sures,I., Lowry,J., Kedes,L.H. (1978) The DNA sequences of sea urchin (S. purpuratus) H2A, H2B, and H3 histone coding and spacer regions. Cell 15 1033-1044

Sures,I., Levy,S., & Kedes,L.H. (1980) Leader sequences of Strongylocentrotus purpuratus histone mRNAs start at a unique heptanucleotide common to all five histone genes. Proc Natl.Sci.USA 77 1265-1269

Sutcliffe,J.G. (1978) pBR322 restriction map marked from the DNA sequence. Accurate DNA size markers upto 4361 nucleotide pairs long. Nucl.Acids Res. 5 2721

Szostak,J.W., Wu,R., (1980) Unequal crossing-over in the ribosomal DNA of Saccharomyces cerevisiae. Nature 284 426-30

Tichter,A., Graur,D. (1989) Nucleic Acid Composition, Codon Usage, and rate of Synonymous substitution in Protein coding genes. J. Mol. Evol. 28 289-294

Trainor,C.,D., Engel,J.,D. (1989) Transcription of the Chicken Histone H5 gene is mediated by Distinct Tissue-specific Elements within the promoter and the 3' Enhancer. Mol. Cell. Biol. 28 286-298

Tsai,S.Y., Tsai,M.J., O'Malley,B.W., (1981) Specific 5' flanking sequences are reqired for faithful initiation of in vitro transcription of the ovalbumin gene. Proc.Nat.Acad.Sci.USA 78 879-883

Tung,L., Harris,G.F., Yager,L.N. Weinberg,E.S. (1980) Sea urchin early and late H4 histone genes bind a specific Transcription factor in a stable preinitiation complex. Mol. Cell. Biol. 9 1476-1487

Turner, P.C., Woodland, H.R. (1982) H3 and H4 histone cDNA sequences from Xenopus: A sequence comparison of H4 genes. Nucl.Acids Res. 10 3769-80

Turner,P.C., Woodland, H.R. (1983) Histone gene number and organization in Xenopus: Xenopus borealis has a homogeneous major cluster. Nucl.Acids Res. 11 971-986

Turner,P.C., Aldridge,T.C., Woodland, H.R., Old,R.W.,(1983) Nucleotide sequences of H3 histone genes from Xenopus laevis. A recently diverged pair of H3 genes and an unusual H3 pseudogene. Nucl.Acids Res. 12 4093

Turner,P.C. Baguzal,R.R.D., Vlad,R.T., Woodland,H.R., (1984) The organization and expression of histone genes from X. borealis. Nucl.Acids Res. 16 3471-3485

Tynowska,J., Finchberg,H., (1973) Chromosome complements of the genus Xenopus. Chromosoma 44 335-342

Van Dongen,W.M.A., de Leaf,L., Moorman,A.F.M., Bentrem,O.M.J., (1981) The organization of histone genes in the genome of Xenopus laevis. Nucl.Acids Res. 9 2297-2311

Van Dongen,W.M.A., Moorman,A.F.M., Bentrem,O.M.J., (1984) DNA organization and expression of Xenopus histone genes. In 'Histone genes and Histone gene expression, ed. G.Stein, J.Stein, N.Marzluff. New York: Wiley

Vinograd,G., (1963) Sedimentation equilibrium in a buoyant density gradient. Meth. in Enz. 6 854-870

Wallis,J.W., Hereford,L., Grunstein,M. (1980) Histone H2B genes of yeast encode two different proteins. Cell 22 799-805

Wallis,J.W., Hereford,L., Grunstein,M., (1980) Histone H2b genes in Yeast encode two different proteins. Cell 22 799-805

Wasylyk,B., Cashon,P., (1981) A T to A base substitution and small deletions in the conalbumin TATA box drastically decreases specific in vitro transcription. Nucl.Acids Res. 9 1813-1824

References

Weinberg,R.S., Overton,G.C., Hendricks,N.B., Hossack,K.M., Chan,L.N. (1978) Histone gene heterogeneity in the sea urchin S. purpuratus. Cold Spring Harbor Symp. Quant. Biol.42 1093-1100

Weinthal,J., Lieberman,D., Hoffman-Liebermann,B., Maxson,R., Maurer,A., Childs,G., Kedes,L.H. (1983) An unusual transposon with long terminal inverted repeats in the sea urchin Strongylocentrotus purpuratus. Quoted in Maxson et al (1983).

Wood,W.B. (1966) Host specificity of DNA produced by E. Coli: Bacterial mutants affected the restriction and modification of DNA. J. Mol. Biol. 16 118

Woodland,H.R. (1980) Histone synthesis during the development of Xenopus. Febs letters 121 1-10

Woodland,H.R. (1982) The translational control phase of early development. Biosciences Rep. 2 471-491

Woodland,H.R., Harrington,J.P., Ballantine,J.E.M., Turner,P.C.,(1984) Are there major developmentally regulated H4 gene classes in Xenopus ? Nucl.Acids Res. 12 4939-4957

Woodland,H.R., Old,R.W., Sturyans,E.A., Ballantine,J.E.M., Aldridge,T.C., Turner,P.C., (1983) The strategy of histone gene expression in the development of Xenopus. Current problems in gene cell differentiation. ed. McLaren,A., Wylie,C.C. Cambridge University press

Wilson,M.C., Melli,M., (1977) Determination of the number of histone genes in human DNA. J. Mol. Biol. 110 511-535

Yamamoto,K.R., Alberts,B.M., Benzinger,R., Lawhorne,L., Treiber,G. (1970) Rapid bacteriophage sedimentation in the presence of polyethylene glycol and its application to large-scale virus purification. Virology 40 734

Zernik,M., Heintz,N., Heims,I., Roeder,R., (1980) Xenopus laevis histone genes: Variant H3 genes are present in different clusters. Cell 22 807-15

Volume 16 Number 8 1988 Nucleic Acids Research

The organisation and expression of histone genes from *Xenopus borealis*

Philip C.Turner*, Edward B.D.Bagenal[1], Marcela T.Vlad[1] and Hugh R.Woodland[1]

Department of Biochemistry, University of Liverpool, PO Box 147, Liverpool L69 3BX and [1]Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK

Received August 20, 1987; Revised and Accepted March 1, 1988

ABSTRACT

We have isolated genomic clones from Xenopus borealis representing 3 different types of histone gene cluster. We show that the major type (H1, H2B, H2A, H4, H3), present at about 60-70 copies per haploid genome (1), is tandemly reiterated with a repeat length of 15 kb. In situ hybridization to mitotic chromosomes shows that the majority of histone genes in Xenopus borealis are at one locus. This locus is on the long arm of one of the small sub-metacentric chromosomes. A minor cluster type with the gene order H1, H3, H4, H2A is present at about 10-15 copies. The genome also contains rare or unique cluster types present at less than 5 copies having other types of organisation. An isolate of this type had the gene order H1, H4, H2B, H2A, H1 (no H3 cloned). Microinjection of all of the clones into Xenopus laevis oocyte nuclei shows that most of the genes present are functional or potentially functional and a number of variant histone proteins have been observed. S1 mapping experiments confirm that the genes of the major cluster are expressed in all tissues and at all developmental stages examined.

INTRODUCTION

The histone genes of animals are arranged in two extreme kinds of way (2). In one, the coding sequences of the 5 histone types (H1, H2A, H2B, H3, H4,) are located close to each other and this quintet is repeated tandemly many times. At the other extreme the genes are arranged in apparently random order (except for H2A and H2B genes tending to be in transcriptionally divergent pairs), some clustered, some dispersed throughout the genome. The closest to the first arrangement is seen in Drosophila melanogaster, where about 100 genes are arranged as tandemly repeated, highly conserved quintets. Two quintet types exist differing only by the presence of a small insertion (3). There are, however, a few dispersed representatives of these genes, called orphons, though there is no evidence that they are expressed (4). At the other extreme are chickens where there are about 6 H5 genes and 8-10 of each core histone gene, mostly scattered on two 50 kb sections of DNA (5), but some variant genes, the H2Ay and H3, are present elsewhere (6,7).

Quite separate from the macro organisation of the histone genes is their individual sequence and pattern of expression. As discussed in several recent reviews (2,8,9,10), histone genes may be restricted in expression to the S-phase, or expressed

independantly of DNA synthesis. In addition they may be restricted to a particular part of the life cycle, like the early genes of sea urchins, or to a particular cell type, like the H5 gene. Only in sea urchins is there a clear correlation between structure and function of their histone genes. These several hundred early genes, expressed between oocyte maturation and gastrulation, are of the conserved, tandemly repeated, quintet type. The genes expressed at later stages are of the disorganised type, like those of chickens and mammals.

We have recently made a preliminary analysis of the histone genes of Xenopus borealis (1). In terms simply of organisation they fit the sea urchin pattern. Genomic Southern blotting shows that about 60% of the genes are present in a single kind of quintet, though it was not clear that they were tandemly repeated. The other genes had some other kind of organisation. Independently it was shown that the same major H4 mRNAs were present throughout the life cycle and in a number of adult cell types (11), though it was not known from which kind of gene they came.

In this paper we establish the detailed organisation of the major gene clusters of X. borealis and establish when and where they are expressed.

MATERIALS AND METHODS

Animals

Xenopus laevis were obtained from the South African Snake Farm, Fish Hoek, South Africa. Xenopus laevis borealis were the first generation raised at Warwick from animals collected in the Kibwezi Forest, Kenya.

Library Construction and Screening

High molecular weight genomic DNA was prepared as previously described from the blood of a single X. borealis female (lane 8, Fig. 3, Ref. 1) and partially digested with Sau 3A. Sucrose gradient fractions containing fragments in the 15-20 kb size range were pooled and ligated to purified Bam HI arms of the vector λ L47.1 (12). Recombinant phage were selected by plating on the P2 lysogenic strain WL 95. Approximately 1 million recombinant phage were screened by the method of Benton and Davis (13) using a mixed H1 and H4 hybridization probe. The H4 probe was the 381 bp Bam HI insert from the H4 cDNA clone pcXlH4W1 (14) and the H1 probe was a 424 bp Msp I fragment from the X. laevis genomic clone XLHW19 containing 266 bp of the coding region and 158 bp of 3' non-coding region (15). An approximately equimolar mixture of these two DNAs was nick-translated with ^{32}P-dCTP and dGTP (16).

Screening of the EMBL 3 partial Sau 3A genomic library was performed using a 0.8 kb Hind III/Sac I fragment from clone λXBH302, marked as probe A in Fig. 1A. DNA was prepared from positive plaques by the rapid plate lysate method (17).

Southern Blot Hybridizations/Gel Analysis/Nuclear Microinjection

The nuclear microinjection procedure, DNA and protein gel analysis and the Southern blot hybridizations were exactly as described in Old et al. (18). The probes used for the Southern blots were as described above for H1 and H4. The H3 probe was a 165 bp Sau 3A/Bam HI fragment from XLHW23 encoding amino acids 74-128 of the H3 protein (19). The H2A probe was a 798 bp Sac I/Xba I fragment from XLHW8 encoding the 3' half of the H2A protein and 349 bp of 3' non-coding region (15,18,20). The H2B probe was a 193 bp Bam HI/Eco RI fragment from XLHW11 encoding amino acids 13-63 of the H2B protein (18,20).

In Situ Hybridization

The hybridization probe was the X. laevis H4 cDNA clone pcXlH4W1 (14) nick-translated (16) using ^{3}H-TTP (40-50Ci/mmole, Amersham). The labelled DNA was phenol-chloroform extracted and ethanol precipitated using E. coli tRNA as carrier. Specific activities varied from 4-8 x 10^4 cpm/µg.

Mitotic chromosomes were made from gut epithelial cells of animals previously injected with colchicine (21).

Prior to hybridization the chromosome preparations were treated with ribonuclease A (100µg/ml in 2 x SSC) for 1 hour at 37°C, washed in 2 x SSC and dehydrated in ethanol. The chromosomal DNA was denatured by submerging the slides in 0.07M NaOH for 3 minutes followed by washes in 70%, 95% and 100% ethanol and air drying. The hybridization reaction contained 40% formamide in 4 x SSC, 0.1M Na_3PO_4, pH7 and a probe concentration of 1-2 x 10^5 cpm/µl. 5-10µl of probe was placed on a mitotic chromosome preparation, a coverslip was added and the edges were sealed with rubber solution. The slides were incubated at 37°C for 19-20 hours. After hybridization and removal of the rubber solution and coverslips the slides were washed at 65°C for 1 hour in 2 x SSC to remove non-specifically bound radioactivity. The dried slides were coated with Kodak NTB_2 diluted 1:1 with H_2O, and left at 5°C for 18-21 days. The autoradiographs were developed in Kodak D19 for 2½ minutes, at room temperature, fixed for 5 minutes, washed in H_2O for 30 minutes and stained with Giemsa.

S1 Nuclease Assays

RNA was made from oocytes, tadpoles and adult tissues as described previously (11). The hybridization probes were all single-stranded M13 subclones of λXbHJ02 uniformly labelled with ^{32}P-dCTP and DNA polymerase I (Klenow fragment) using the M13 universal primer (22).

Hybridization reactions using the amounts of RNA given in the figure legends were carried out essentially according to Berk and Sharp (23) and gel analysis was as performed in our earlier work (11,24).

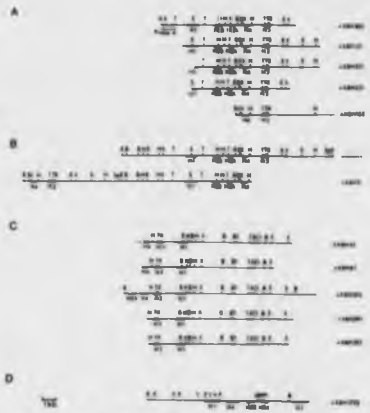


Figure 1 Genomic Clones Containing Histone Genes from X. borealis. (A) Organisation of λ L47.1 clones of the major histone cluster type. In λXBH302, the position of a DNA fragment used as a probe for the EMBL3 library and for the experiment of Fig. 2 is marked as Probe A. (B) Organisation of EMBL3 clones of the major histone cluster type. (C) Organisation of λ L47.1 clones of the minor cluster type. (D) A rare/unique histone cluster in λL47.1. The restriction enzymes used have been abbreviated as follows: B, Bam HI; E, Eco RI; H, Hind III; X, Xba I; S, Sal I; T, Sst I/Sac I; M, Sma I; K, Kpn I; O, Xho I; V, Pvu II; P, Pst I; Sp, Sph I. Not all enzymes listed have been used to map all clones. Clone λXBH52 is thought to contain a double insert.

RESULTS

Organisation of X. borealis Histone Clones

On initial screening of the λ L47.1 genomic library with a mixed H1/H4 probe, 11 independent clones were isolated. Restriction mapping and subsequent probing of Southern blots with specific gene probes as detailed in materials and methods and DNA sequence analysis (not shown) gave rise to the data shown in Fig. 1.

A group of related clones were isolated that had exactly the organisation predicted from our earlier genomic mapping studies (1) and are therefore

representatives of the major cluster type in X. borealis (Fig. 1A). The gene order of this major cluster is H1. H2B. H2A, H4, H3 and from sequencing data (not shown) the H2B and H4 genes have the opposite polarity to the other 3 genes. We have previously shown that the copy number of this major cluster type is 60-70 copies per haploid genome (1).

A second group of clones were obtained (Fig. 1C) that contain a cluster of histone genes that are organised differently to the major cluster type. The gene order of this minor type is H2A, H4, H3, H1. Several of the restriction fragments that make up this minor cluster were observed as minor bands of hybridization on genomic Southern blots in our earlier studies (1). From the relative intensities of the bands we can now estimate that this minor cluster type is present at approximately 10-15 copies per haploid genome. None of the clones isolated contains an H2B gene as judged by mapping and sequencing experiments (not shown), but the position of the four coding regions near one end of each isolate makes it likely that this cluster type contains at least one copy of each of the 5 histone classes.

Two independent isolates of a third cluster type were also obtained. Only one of these was analysed in detail as initial mapping showed them to be very similar (Fig. 1D). This cluster has the following organisation: H1, H4, H2B, H2A, H1 (no H3 gene present) with the H2A and H2B genes having opposite polarity. Again restriction fragments that comprise this clone were observed in our earlier genomic mapping experiments and we can now conclude that this cluster type has a copy number of less than 5 per haploid genome. This gene organisation is unusual in that 2 H1 genes are present but no H3 gene has been cloned and thus this cluster type cannot be a simple quintet. Some genomic clones isolated from X. laevis do not have a quintet structure in that the cluster contains 2 H4 genes (26).

We note that the gene-containing, restriction fragments for the major and minor cluster types generated by Eco RI cleavage of X. borealis genomic DNA are not easily resolved on gels but are clearly separable when Bam HI is used. This highlights the pitfalls of attempting to map and analyse repetitive gene families using genomic blots and it may account for some of the discrepancies in the literature regarding histone gene organisation in X. laevis (25,18,26).

Tandem Repetition of the Major Cluster

To try to establish whether the major cluster is tandemly linked, another X. borealis genomic library, constructed in EMBL 3 (a gift of C. Wilson) was screened using a small fragment from the left-hand end of clone λXBH302 (probe A, Fig. 1A).

The restriction maps of two clones from this library, λXBH9 and 11 (Fig. 1B), suggest that at least one pair of major clusters are contiguous. Clone λXBH9 contains a single Sph I site in the intergenic region which allowed us to ask if the majority of the H4 genes in the genome were present in this organisation. If most of the major repeats

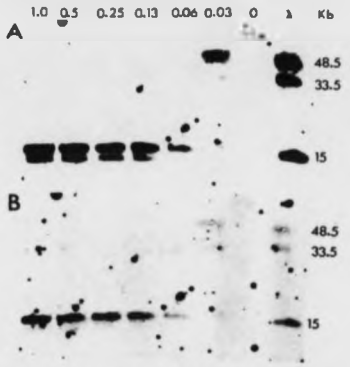


Figure 2 Southern Blot of Totally and Partially Sph I - digested Genomic DNA. Panel A was probed with the H4 cDNA clone, pcXIH4W1. Panel B was probed with the HindIII/SalI intergenic probe, specific to the major histone gene cluster, marked as probe A in Fig. 1/A. 2 µg of genomic DNA was digested for 1h at 37°C at from 1.0-0 units of Sph I/µg DNA, as shown above the lanes. Mixed radioactive markers of digested and undigested λDNA were also run, their sizes being indicated on the right. The DNA was electrophoresed on a 0.4% agarose gel for 24 hours which was acid-treated before blotting.

are present as a tandemly repeated 15 kb unit, then an Sph I digest of genomic DNA, probed with an H4 probe would be expected to show a single strong band of 15kb, as well as some very minor bands. The actual experiment is presented in Fig. 2A and shows a predominant 15 kb band. Partial digests also generate a 30kb band, which is entirely consistent with tandem repitition. Larger molecules are not resolved on the gel. There is also a more minor band of 14.2kb in the limit digest. This is probably, at least partly, a minor variant of the major cluster, since it hybridizes, as seen in Fig. 2B, with a major cluster-specific probe, probe A of Fig. 1A. It is not the minor cluster, since a probe from between the H3 and H4 genes of λXBH61 reveals a band slightly smaller than 14kb (data not shown); this is sufficiently close to the lower H4 band of Fig. 2A for it to be subsumed into it in the H4 probing experiment.

Thus the intercluster organisation seen in clone λXBH9 of Fig. 1B seems the

Figure 3 In situ Hybridization Experiments. (A) A mitotic complement of X. borealis after in situ hybridization with ^{3}H-labelled, H4 coding sequences. One pair of chromosomes is labelled at the end of the long arm (arrows). Because of similar length and centromeric position the labelled pair could occupy any one of the positions 14 to 17. (B) Several examples (from the same animal as in A) of terminally labelled chromosomes after in situ hybridization with the same histone probe as in A. (C) A partial mitotic complement of X. borealis after in situ hybridization as above. The arrows indicate the two labelled chromosomes. The interphase nucleus (top left) shows a typical labelling pattern (see text).

predominant arrangement in the genome, and the majority of the major clusters must therefore be tandemly linked with a repeat length of 13 kb. This value fits well with our previous genomic mapping data (1).

Chromosome Location

We reasoned that if the 60-70 copies of the major cluster were tandemly linked

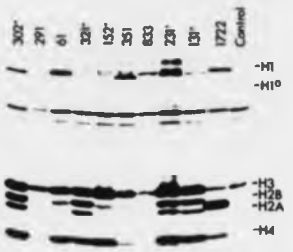


Figure 4 Analysis of histones made by the genomic clones analyzed by SDS gel electrophoresis. The genomic clones were injected into the nuclei of X. laevis oocytes, which were incubated overnight. Approximately 20 oocytes from each group were then incubated overnight in 40 µl ^{3}H-lysine (5 mCi/ml). The nuclei were dissected from these oocytes, selectively extracted as previously described (18) and one third of the sample was run on a 40 cm SDS/18% acrylamide gel. The positions of stained X. borealis histones are indicated. The clones are identified above the tracks, representatives of the major cluster being marked with an asterisk.

then it would be possible to determine their chromosome location. When H4 coding sequences were hybridized to mitotic chromosomes of X. borealis one locus was prominently labelled, at the end of the long arm of a small sub-metacentric chromosome (Fig. 3A and B). Because many chromosomes of the X. borealis complement have similar length and centromere position, precise identification of the labelled chromosome pair is not possible without quinacrine banding. Although Fig. 3 shows the labelled chromosome pair at the position I6, it can be any one from I4 to I7.

The interphase nuclei also showed only 1 or 2 labelled regions (Fig. 3C, top left) which is suggestive of a single locus containing the majority of the histone sequences. However, because the total number of silver grains in the autoradiograph is low, scattered histone sequences would not be detected in our preparations.

Expression of the Genes Following Microinjection into X. laevis Oocytes

To establish whether the genomic clones encoded functional histone genes that corresponded to the histone proteins extracted from X. borealis tissues the linear DNAs were microinjected into X. laevis oocyte nuclei. This experiment would also give a preliminary indication of the degree of diversity of the histone genes which is valuable since analysing them all in detail would be a major undertaking. After overnight incubation the oocytes were labelled with ^{3}H-lysine and ^{35}S-methionine for

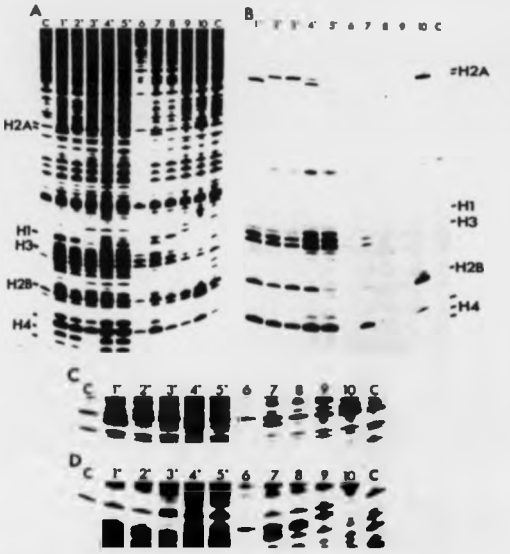


Figure 3 Analysis of histones made by the genomic clones analyzed on Triton/Acetic acid/Urea gels. One third of the same samples as in Fig. 4 was run on a 40 cm triton/acetic acid/urea gel. A shows a 2 month exposure. That in B is 6 hours. C and D are enlargements of the H2A and H1 regions of the longer exposure. The clones injected were as follows: 1, λXBH302; 2, λXBH431; 3, λXBH321; 4, λXBH231; 5, λXBH452; 6, λXBH291; 7, λXBH461; 8, λXBH435; 9, λXBH433; 10, λXBH4722; C, control oocytes. The positions of marker X. borealis erythrocyte histones are marked. Representatives of the major cluster are marked with an asterisk.

24 hours and the extracted proteins were analysed on both SDS and triton/urea polyacrylamide gels (Fig. 4 and Fig. 5 respectively).

All the clones of the major cluster shown in Fig. 1A made H2B, H3 and H4, except clone λXBH152 which contains only H3 and H4 sequences and in every case these proteins co-migrated with the major types in bulk histone preparations. All except clone λXBH152 also made H2A, but two clones, λXBH231 and λXBH302 encoded H2A proteins that migrated faster than normal on triton/urea gels. In view of the extreme nucleotide sequence conservation of representatives of the major cluster (unpublished observations), it seems probable that this change in mobility is due to a single amino acid substitution in each case. In λXBH302 threonine replaces the more usual alanine at position 33 of the H2A as a result of a single base change.

Of the major cluster clones, λXBH302 and λXBH131 expressed H1 proteins probably because the H1 sequences in λXBH231 and λXBH321 are terminal. The mobility of the expressed major cluster H1 gene products in clones λXBH131 and λXBH302 is similar to that of the X. laevis H1C type in migrating faster than the X. laevis H1A or H1B type on SDS gels (15). Therefore expression gives a picture of very limited heterogeneity in these genes, applying only to the H2A genes.

The group of clones representing the minor cluster (Fig. 1C) all expressed H3 protein with a similar mobility to normal/bulk H3. Though they all also expressed H1 protein, the mobilities of these H1 proteins were quite variable. We observed that the single H1 band on SDS gels was resolved into 2 bands on triton/urea gels presumably as a result of a partial modification. In general the H1 proteins encoded by the minor cluster clones migrated less quickly on both SDS and triton/urea gels than the major H1 type. This may be consistent with the view that X. borealis contains H1 protein types related to the X. laevis H1A and H1B types. However, the degree of variability is quite large since none of the 4 representatives of the minor cluster type seem to contain identical H1 genes when analysed by mobility on 2 gel systems. This contrasts with the identical behaviour of the H1 proteins in the major cluster clones. We have previously noted for a pair of very similar H1C genes in X. laevis (15) that H1 gene variability is not restricted to amino acid substitutions but includes significant insertions and deletions.

Clone λXBH61 seemed to generate two labelled protein bands in the H1 position on SDS gels (and 3 on triton/urea). Further analysis of this clone reveals only one H1 coding region present which must therefore give rise to these two bands. The most likely explanation is a post-translational modification of some of the newly synthesised protein. Nucleotide sequence analysis will hopefully explain this observation.

Clone λXBH722, containing a rare cluster type expressed a normal H4 protein but variant H1, H2A and H2B proteins. Since only one H1 protein band is visible by gel analysis either the two H1 genes present are identical or one is not expressed. If the

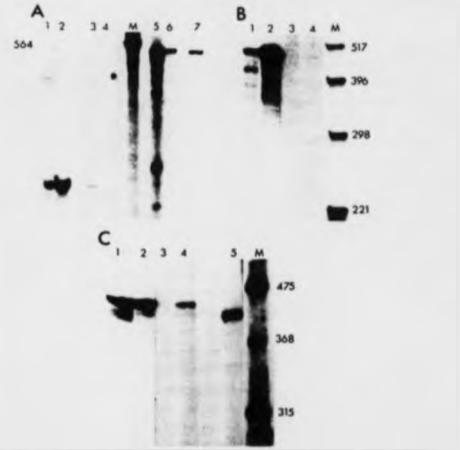


Figure 6. S1 nuclease analysis of histone gene transcripts from the major cluster clone λ XH-HIII in X. borealis tissues. (A) Analysis of H2B (lanes 1-4) and H2A expression (lanes 5-7). For H2B we used an M13 subclone from a Hind III site within the gene, through the 3' end of the gene. Protection to the 3' end of the mRNA would give 206n. For H2A we used an M13 subclone spanning the entire transcribed region from an upstream Hind III site to a downstream Bam HI site. Protection of the entire mRNA would give 501 n. Lanes 1 and 5, 63 µg ovary RNA; 2 and 6, 10 µg tadpole RNA; 3, 21 µg liver cell RNA; 4 and 7, 2.5 µg heart cell RNA; M is λ Hind III marker. (B) Analysis of H4 expression using an M13 subclone from a Sal I site within H4 through the 3' end of the gene. Protection to the 3' end of the mRNA would give 469n. Lane 1, 54 µg ovary RNA; 2, 8 µg tadpole RNA; 3, 12.5 µg lung cell RNA; 4, 10 µg heart cell RNA; M is pAT153 Hinf I marker. (C) Analysis of H3 expression using an M13 subclone from a Bam HI site through the 5' end of the gene. Protection to the 5' end of the mRNA would give 432n. Lane 1, 63 µg ovary RNA; 2, 10 µg tadpole RNA; 3, 21 µg liver cell RNA; 4, 12 µg lung cell RNA; 5, 42 µg X. laevis ovary RNA; M is pBR322 Taq I marker.

latter is true, this would probably be the H4 gene near the end of the clone which may lack some essential sequence. The very anomalously migrating H2A and H2B proteins result from a 2 amino acid deletion (ala-pro) relative to the major H2B at position 10/11

and 2 substitutions of threonine in H2A relative to the major H2A at positions 124 and 126. (Z. Freeman unpublished results)

Developmental and Tissue Specific Expression of the major quintet genes

We have analyzed a number of tissues and developmental stages to see if the major cluster histone genes showed developmental regulation. We previously used primer extension sequencing of H4 mRNAs to show that a single family of transcripts was present throughout development in X. borealis (11). This set of sequences differed by a single base substitution in different animals. We have now shown that the H4 genes of the major quintet clones have an identical sequence to this some having one variant and some the other (unpublished observations). Thus these genes must be expressed throughout development so we have not analysed H4 expression further.

For the other genes in λXBH302 we have prepared uniformly labelled single stranded DNA probes from M13 subclones (see Methods). These were utilized in S_1 nuclease analyses, and the products were run on acrylamide gels (Fig. 6). In every case the longest bands obtained were of the length predicted from the sequence of the gene. The analysis included total RNA preparations from ovary stage 67 swimming tadpoles adult liver and primary cell cultures from adult heart and lung. All generated the fragments expected of the homologous transcripts although the signal was lower from adult than from embryonic tissues. We can conclude that these genes are expressed throughout development. In the case of H3, below the fully protected band is seen a weaker band corresponding to protection to the translation start site. As expected this is the largest band seen in X. laevis RNA. The fact that the longest band is always strongest in the X. borealis tissues suggests that H3 genes of the major cluster type always make up the bulk of the H3 mRNA, at least in the tissues tested. For each gene there are a range of more minor, smaller bands which will certainly represent transcripts with sequence differences (noteably for H1), partially degraded RNA molecules, internal cuts in the RNA/DNA hybrids and probe radiolysis.

DISCUSSION

In this paper we have proven that our earlier map of the major histone cluster in X. borealis based on genomic Southern blots is correct being H1, H2B, H2A, H4, H3. In addition we have determined that the H2B and H4 genes have the opposite polarity to the other 3 genes. The repeat unit length is 15 kb and the chromosomal location of this tandemly repeated major cluster is near the end of one of the shortest sub metacentric chromosomes. The gross organisation of the major cluster is similar to that in Drosophila (3,4), both in terms of gene order and polarity but not in repeat length, yet different from sea urchin (27,28), newt (29), X. laevis (18 24 30) and X. tropicalis (31). This is not very likely to occur by chance. However, considering the relatively small

type (to which the X. borealis minor type is related) is tandemly reiterated we are hopeful that by choosing a suitable cluster specific probe we can discover more about the chromosomal organisation of the minor types in X. borealis

The microinjection experiments presented here illustrate an unexpected degree of protein diversity among X. borealis histone proteins In the case of the most anomalously migrating H2A and H2B proteins in λXBH4722 the preliminary sequence data (Z Fraenkel unpublished results) indicate that there is a two amino substitution in the H2A and a two amino acid deletion in the H2B The protein gels indicate that at least 3 different H1 species are present in our clones This is not entirely surprising in view of the H1 gene diversity we have found in X. laevis (15) and a detailed sequence analysis is underway.

Our S_1 nuclease assays did not detect regulation in the developmental or tissue specific expression of genes in the major cluster

ACKNOWLEDGEMENTS

We would like to thank Julie McGreavey for typing the manuscript C Klass for technical assistance R.W. Old A Mohammed T.C Aldridge and Z Fraenkel for providing unpublished data and C Wilson for his generous gift of the EMBL3 genomic library This work was supported by an MRC grant to H.R.W

*To whom correspondence should be addressed

REFERENCES

1 Turner, P.C. and Woodland. H.R. (1983) Nucl. Acids Res. 11, 971-984.
2. Stein, G.S., Stein. J.L. and Marzluff. W.F. (1984) Histone Genes. John Wiley and Sons, New York.
3. Lifton, R.P., Goldberg, M.L., Karp. R.W. and Hogness. D.S. (1978) Cold Spring Harbor Symp. Quant. Biol 42. 1047-1051.
4. Childs, G., Maxson, R., Cohn. R.H. and Kedes. L. (1981) Cell 23. 651-663.
5. Grandy, O.K. and Dodgson, J.B. (1987) Nucl. Acids Res. 15. 1063-1080.
6. Harvey, R.P., Whiting, J.A., Coles, L.S., Krieg. P.A. and Wells. J.R.E. (1983) Proc. Natl. Acad. Sci. USA. 80, 2819-2823.
7 Krieg, P.A., Robins. A.J., Gait. M.J., Titmas. R.C. and Wells. J.R.E. (1982). Nucl. Acids Res. 10. 1495 1502.
8 Old. R.W. and Woodland. H.R. (1984) Cell 38 624-625.
9 Hentschel. C.C. and Birnstiel M.L (1981) Cell 25. 301-313.
10. Maxson, R., Cohn. R., Kedes. L. and Mohun. T. (1983) Ann. Rev. Genet. 17 239-277.
11 Woodland, H.R., Warmington. J.R., Ballantine. J.E.M. and Turner. P.C. (1984) Nucl. Acids Res. 12, 4939-4958.
12. Losman, W.A. and Brammar. W.J. (1980) Gene 10. 249-254.
13. Benton, W.D. and Davis. R.W. (1977) Science 196. 180-182.
14. Turner, P.C. and Woodland. H.R. (1982) Nucl. Acids Res. 10. 3769-3780.
15. Turner, P.C., Aldridge. T.C., Woodland. H.R. and Old. R.W. (1983) Nucl. Acids Res. 11. 4093-4107.

16. Rigby, P.W.J., Diechmann, M., Rhodes, C. and Berg, P. (1977). J. Mol. Biol. 113 237-51
17. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A laboratory manual. Cold Spring Harbor Laboratory Press. New York
18 Old, R.W., Woodland, H.R., Ballantine, J.E.M., Aldridge, T.C., Newton, C.A., Baine, W.A. and Turner, P.C. (1982) Nucl. Acids Res 10 7561-7580.
19 Mohammad A., Aldridge, T.C. and Old, R.W. In preparation.
20. Aldridge, T.C. and Old R.W. unpublished observations.
21. Macgregor, H.C. and Andrews, C. (1977). Chromosoma (Berl.) 63 109 126.
22. Hu N and Messing, J. (1982) Gene 17, 171-178.
23. Berk, A.J. and Sharp, P.A. (1977) Cell 12 721-732.
24. Ballantine, J.E.M. and Woodland, H.R. (1985) FEBS Lett. 121, 1-7.
25. Van Dongen, W., De Leaf, L., Zaal, R., Moorman, A. and Destree, O. (1981) Nucl. Acids Res. 9, 2297-2311.
26. Destree, O.H.J., Bendig, M.M., De Leaf, R.T.M. and Koster J.G. (1984) Biochim Biophys. Acta 782. 132-141
27. Gross, K., Schaffner, W., Telford, J. and Birnstiel, M.L. (1976) Cell 8. 479-487.
28. Wu M., Holmes, D.S., Davidson, N., Cohn, R.H. and Kedes, L.H. (1976) Cell 9, 163-171.
29 Stephenson, E.C., Erba, H.P. and Gall, J.G. (1981) Cell 24 639-647.
30. Zernik, M., Heintz, N., Baime, I. and Roeder, R.G. (1980) Cell 22. 807-822.
31. Ruberti, I., Fragapane, P., Pierandrei-Amaldi, P., Beccari, E., Amaldi, F. and Bozzoni, I. (1982) Nucl. Acids Res. 10 7543-7559.
32. Buesinger, M., Rusconi, S. and Birnstiel, M.L. (1982) EMBO J. 1, 27-33.

TITLE A STUDY OF THE HISTONE GENES OF *XENOPUS BOREALIS*

AUTHOR Edward Beauchamp Dudliegh Bagenal

INSTITUTION and DATE University of Warwick 1990